STIC Search Report
Biotech-Chem Library

USPTO

TO: James Schultz
Location: REM-2D18/2C18
Art Unit: 1635
Tuesday, March 23, 2004
Case Serial Number: 10/054313

From: Paul Schulwitz
Location: Biotech-Chem Library
REM-1A65
Phone: (571)272-2527

paul.schulwitz@uspto.gov

Search Notes

Examiner Schultz,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Paul Schulwitz
Technical Information Specialist
STIC Biotech/Chem Library
(571)272-2527

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: March 22, 2004, 21:38:16 ; Search time 2831 Seconds
(without alignments)
4378.703 Million cell updates/sec

Title: US-10-054-313-1
Perfect score: 1546
Sequence: 1 MSWLLPLAHRVALAALPCRR.........FIGNEEADRLAREGAKQSED 286

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop   6.0 , Delext 7.0

Searched: 3470272 seqs, 2167151995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10054313/runat_22032004_140415_28070/app_query.fasta_1.455
-Db=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
-USER=US10054313_@CGN_1_12372_@runat_22032004_140415_28070 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl :
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1526	98.7	1158	9	BC002973	BC002973 Homo sapi
	2	1526	98.7	1168	9	AF039652	AF039652 Homo sapi
	3	1526	98.7	1604	9	AK075490	AK075490 Homo sapi
	4	1523	98.5	1150	6	BD223780	BD223780 Human RNA
	5	1522	98.4	1113	9	AF048994	AF048994 Homo sapi
	6	1522	98.4	1147	9	AF048995	AF048995 Homo sapi
	7	1518	98.2	861	9	HSAJ4117	AJ224117 Homo sapi
	8	1457.5	94.3	2120	9	AK096913	AK096913 Homo sapi
	9	1197.5	77.5	1430	10	AF048993	AF048993 Mus muscu
	10	1193.5	77.2	1409	10	BC019411	BC019411 Mus muscu
	11	1174	75.9	178965	3	AC098850	AC098850 Homo sapi
c	12	1174	75.9	192337	3	AC022596	AC022596 Homo sapi
c	13	1165	75.4	211403	9	AC126352	AC126352 Homo sapi
	14	1150.5	74.4	207418	9	AC090774	AC090774 Homo sapi
	15	1145.5	74.1	176040	9	AC107926	AC107926 Homo sapi
c	16	1114	72.1	188017	9	AC146897	AC146897 Gorilla g
	17	1016	65.7	131239	9	HS667H12	AL035414 Human DNA
	18	956	61.8	2024	9	AY063502	AY063502 Cercopith
	19	923	59.7	1992	9	AY063503	AY063503 Macaca as
	20	882	57.1	1293	5	CHKESTFL25	D26340 Gallus gall
	21	816	52.8	2165	5	AF448495	AF448495 Callithri
c	22	755.5	48.9	1911	9	AK057473	AK057473 Homo sapi
c	23	755.5	48.9	3459	9	BC041488	BC041488 Homo sapi
	24	423.5	27.4	1153	3	AY089374	AY089374 Drosophil
	25	375	24.3	1460	3	AF032921	AF032921 Drosophil
	26	373.5	24.2	1117	3	TBU74470	U74470 Trypanosoma
c	27	373.5	24.2	141775	2	AC105378	AC105378 Trypanoso
c	28	356.5	23.1	37983	2	AC020394	AC020394 Drosophil
c	29	356.5	23.1	85651	2	AC005463	AC005463 Drosophil
c	30	356.5	23.1	86889	3	AC005448	AC005448 Drosophil
	31	356.5	23.1	139410	3	AC007084	AC007084 Drosophil
c	32	356.5	23.1	254023	3	AE003839	AE003839 Drosophil
	33	331.5	21.4	34828	3	U41994	U41994 Caenorhabdi
c	34	312	20.2	166143	2	AL356960	AL356960 Homo sapi
	35	303.5	19.6	2277	2	TRFRNH1A	L8916 Crithidia f
c	36	302	19.5	43680	9	AC145667	AC145667 Homo sapi
c	37	300.5	19.4	56083	9	AC108488	AC108488 Homo sapi
	38	292.5	18.9	38360	2	AC145674	AC145674 Homo sapi
c	39	286	18.5	206332	10	AC136986	AC136986 Mus muscu
	40	276	17.9	39512	8	SPBC336	AL121815 S.pombe c
c	41	272	17.6	1557	8	AF048992	AF048992 Schizosac
c	42	272	17.6	119211	6	AR408761	AR408761 Sequence
c	43	272	17.6	119211	6	AX067465	AX067465 Sequence
	44	268	17.3	231	6	AR422883	AR422883 Sequence
	45	268	17.3	231	6	BD118436	BD118436 EST and e

ALIGNMENTS

RESULT 1

```
LOCUS       BC002973                1158 bp    mRNA    linear   PRI 12-NOV-2003
DEFINITION  Homo sapiens ribonuclease H1, mRNA (cDNA clone MGC:2019
            IMAGE:3537074), complete cds.
ACCESSION   BC002973
VERSION     BC002973.1  GI:12804228
KEYWORDS    MGC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1158)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22388257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1158)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-FEB-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Genome Sequence Centre,
            BC Cancer Agency, Vancouver, BC, Canada
            info@bcgsc.bc.ca
            Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
            Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
            Letticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
            Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
            Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
            Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
            Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
            George Yang, Scott Zuyderduyn, Marco Marra.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 8 Row: 1 Column: 11
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 21359815.
FEATURES             Location/Qualifiers
     source          1..1158
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="MGC:2019 IMAGE:3537074"
                     /tissue_type="Lung, small cell carcinoma"
                     /clone_lib="NIH_MGC_7"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
     gene            1..1158
                     /gene="RNASEH1"
                     /note="synonym: H1RNA"
                     /db_xref="LocusID:246243"
                     /db_xref="MIM:604123"
     CDS             71..931
                     /codon_start=1
                     /product="ribonuclease H1"
                     /protein_id="AAH02973.1"
                     /db_xref="GI:12804229"
                     /db_xref="LocusID:246243"
                     /translation="MSWLLFLAHRVALAALPCRRGSRGFGMFYAVRRGRKTGVFLTWN
                     ECRAQVDRFPAARFKKFATEDEAWAFVRKSAPEVSGLEHENQHGQESEAKASKRLREP
                     LDGDGHESAEPYAKHMKPSVEPAPPVSRDTFSYMGDFVVVTDGCCSSNGRRRPRAGI
                     GVYWGPGHPLNVGIRLPGRQTNQRAEIHAACKAIEQAKTQNINKLVLYTDSMFTINGI
                     TNWVQGWKKNGWKTSAGKEVINKEDFVALERUTQGMDIQWMFVPGHSGPIGNEEADRL
                     AREGAKQSED"
     misc_feature    488..913
                     /note="rnaseH; Region: RNase H. RNase H digests the RNA
                     strand of an RNA/DNA hybrid. Important enzyme in
                     retroviral replication cycle, and often found as a domain
                     associated with reverse transcriptases. Structure is a
                     mixed alpha+beta fold with three a/b/a layers"
                     /db_xref="CDD:pfam00075"
ORIGIN
```

Alignment Scores:

Pred. No.:   1158   Length: 1158
Score: 1526.00   Matches: 282
Percent Similarity: 99.30%   Conservative: 2
Best Local Similarity: 98.60%   Mismatches: 2
Query Match: 98.71%   Indels: 0
DB: Gaps: 0

US-10-054-313-1 (1-286) x BC002973 (1-1158)

```
Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuAlaAlaLeuProCysArgArg  20
Db   71 ATGAGCTGGCTTCGTTCCTGGCCCACAGAGTCGCTTGCCCTGCCCTGCCGCCCGC 130

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe  40
Db  131 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTAGCAAGACCGGGGTCTT 190

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys  60
Db  191 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCCGGTTTCCTGCTGCCAGATTTAAGAAG 250

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer  80
Db  251 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAGCCCGGAAGTTTCA 310

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysProGlyLeuArg 100
Db  311 GAAGGGCATGAAAATCAACATGAACATGGAGAAGATCGGAGGCGAAAGCCAGCAAGCGACTCCGT 370

Qy  101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db  371 GAGCCACTGGATGGAGATGGACATGAAAGGCCAGACCGTATGCAAAGCACATGAAGCCG 430

Qy  121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db  431 AGCGTGGAGCCGGCGCCTCCAGTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC 490

Qy  141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgArgProArgAlaGlyIle 160
Db  491 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGGAGCAGGAGAATC 550

Qy  161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db  551 GGCGTTTACTGGGGGCCAGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGGCAG 610
```

```
Query Match:        98.71%    Indels:    0
DB:                    9       Gaps:      0

US-10-054-313-1 (1-286) x AF039652 (1-1168)

Qy      1  MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg   20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     82  ATGAGCTGGCTCTGTTCCTGGCCCACAGAGTCGCCCTAGCCGCCCTTGCCCTGCCGCGCC  141

Qy     21  GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyLysThrGlyValPhe   40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    142  GGCTCTCGGGGGTTCGGGATGTTCTATGCCGTGAGGGGCCGCAAGACCGGGGTCTTT  201

Qy     41  LeuThrTrpAsnGlyGlnArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys   60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    202  CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG  261

Qy     61  PheAlaThrGluAspGluAlaTrpAlaPheValLysSerAlaSerProGluValSer   80
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    262  TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAAGAATCTGCAAGCCCGGAAGTTTCA  321

Qy     81  GlyGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg  100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    322  GAAGGGCATGGAAAATCAACATGGACAAGAATCGGAGGCGAAAGCCAGCCAGCGACTCCGT  381

Qy    101  GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    382  GAGCCACTGGATGGAGATGGACATGAAAGCGCAGAAGCCCTATGCAAAGCACATGAAGCCG  441

Qy    121  SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    442  AGCGTGGAGCCCGCGCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC  501

Qy    141  ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgArgProArgAlaGlyIle  160
           |||||||||||||||||||||||||||||::|:|||||||||||||||||||||||||||
Db    502  GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGGAGCAGGAATC  561

Qy    161  GlyValTyrTrpGlyTyrProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    562  GGCGTTTACTGGGGCTACCCAGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCCAG  621

Qy    181  ThrAsnGlnArgAlaGluHisAlaAlaCysLysAlaIleGluGlnAlaLysLysThrGln  200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    622  ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAGACAAAGACTCAA  681

Qy    201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    682  AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC  741

Qy    221  TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyLysValIleLeuAsn  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    742  TGGGTTCAAGGTTGGAAGAAAATGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC  801

Qy    241  LysGluAspPheValAlaLeuAlaLeuThrArgLeuThrGlnGlyMetAspIleGlnTrpMetHis  260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    802  AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT  861

Qy    261  ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaLysArgGluGlyTyr  280
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    862  GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAAGAGAGGA  921

Qy    281  AlaLysGlnSerGluAsp  286
           |||||||||||||||||
Db    922  GCTAAACAATCGGAAGAC  939

RESULT 3
AK075490
LOCUS       AK075490                1604 bp    mRNA     linear    PRI 03-SEP-2002
DEFINITION  Homo sapiens cDNA PSEC0185 fis, clone OVARC1002091, highly similar
            to Homo sapiens ribonuclease H type II mRNA.
ACCESSION   AK075490
VERSION     AK075490.1  GI:22761667
KEYWORDS    oligo capping; fis (full insert sequence).
```

```
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Ota,T., Nishikawa,T., Suzuki,Y., Kawai-Hio,Y., Hayashi,K.,
            Ishii,S., Saito,K., Yamamoto,J., Wakamatsu,A., Nagai,T.,
            Nakamura,Y., Nagahari,K., Sugano,S. and Isogai,T.
  TITLE     HRI human cDNA sequencing project
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 1604)
  AUTHORS   Isogai,T. and Yamamoto,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAR-2002) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT     HRI human cDNA sequencing project; cDNA 5'- & 3'-end one pass
            sequencing, clone selection and full insert sequencing: Helix
            Research Institute (supported by Japan Key Technology Center etc.);
            cDNA library construction: Institute of Medical Science, University
            of Tokyo, Laboratory of Genome Structure, Human Genome Center.
FEATURES             Location/Qualifiers
     source          1..1604
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="OVARC1002091"
                     /tissue_type="ovary, tumor tissue"
                     /clone_lib="OVARC1"
                     /note="cloning vector: pME18SFL3"

ORIGIN

Alignment Scores:
Pred. No.:
Score:                    2.46e-101
Percent Similarity:       1526.00          Length:          1604
Best Local Similarity:    99.30%           Matches:         282
Query Match:              98.60%           Conservative:    2
DB:                       98.71%           Mismatches:      2
                                           Indels:          0
                                           Gaps:            0

US-10-054-313-1 (1-286) x AK075490 (1-1604)

Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuAlaLeuProCysArgArg   20
Db   61 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTGGCCCTTGCCGCCCTGCCGTGCCGGCGC  120

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyThrGlyValPhe   40
Db  121 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGAAGACCGGGGTCTTT  180

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAlaAspArgPheAlaAlaArgPheLysLys   60
Db  181 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCAGATTAAGAAG  240

Qy   61 PheAlaThrGluAspGlnAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
Db  241 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCCGGAAGTTTCA  300

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGlnGluSerGlyGlnLeuSerProGlyLysArgLeuArg  100
Db  301 GAAGGGCATGAAAATCAACATGAACAAGAATCGGAGGGCAGGCCGAAAGCCAGCCACTCCGT  360

Qy  101 GluProLeuAspGlyAspGlyPheHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
Db  361 GAGCCACTGGACGGAGATGGACATGAAAGCGCAGCCGTATGCAAAGCACATGAAGCCG  420

Qy  121 SerValGluProAlaProProValSerArgSerArgAspThrProSerTyrMetGlyAspPheVal  140
Db  421 AGCGTGGAGCCGGCGCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC  480

Qy  141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgProArgProAlaGlyIle  160
Db  481 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCGAGCAGAATC  540

Qy  161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgProGlyArgGln  180
Db  541 GGGCGTTTACTGGGGCGGGCCGGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGGCAG  600

Qy  181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysLysAlaIleGluGlnAlaLysThrGln  200
Db  601 ACAAACCAAAGAGCAGGAAATTCATGCAGCCTGCAAAGCCATTGAACAGCAAAGACTCAA  660

Qy  201 AsnIleAsnLysAsnValLeuTyrThrArgSerMetTheThrIleAsnGlyIleThrAsn  220
Db  661 AACATCAATAAAACTGGTTCTGTATACAGCAGTATGTTCACGATAAATGGTATAACTAAC  720

Qy  221 TrpValGlnGlyTrpLysLysAsnGlyTyrTrpLysSerAlaGlyTrpLysGluValIleAsn  240
Db  721 TGGGTTCAAGGTTGGAAGAAAAATGGGTACTGGAAGAGCAGGGAAAGAAGGTGATCAAC  780

Qy  241 LysGluAspPheValAlaLeuGluLeuThrGlnGlyMetAspIleGlnTrpMetHis  260
Db  781 AAAGAGGACTTTGTGGCACTGGAGCTTACCCAGGGGATGGACATTCAGTGGATGCAT  840

Qy  261 ValProGlyHisSerGlyPheIleGluGlyAlaArgLeuAlaArgGluGly  280
Db  841 GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGCACAGATTAGCCAGAGAAGGA  900

Qy  281 AlaLysGlnSerGluAsp  286
Db  901 GCTAAACAATCGGAAGAC  918

RESULT 4
BD223780
LOCUS       BD223780                1150 bp    DNA     linear   PAT 17-JUL-2003
DEFINITION  Human RNA-associated proteins.
ACCESSION   BD223780
VERSION     BD223780.1  GI:33033550
KEYWORDS    JP 2002523045-A/16.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1150)
  AUTHORS   Hillman,J.L., Yue,H., Tang,Y.T., Corley,N.C., Guegler,K.J.,
            Gorgone,G.A., Patterson,C., Baughn,M.R., Lal,P., Bandman,O.,
            Reddy,R., Azimzai,Y., Shih,L.L., Yang,J. and Lu,D.A.M.
  TITLE     Human RNA-associated proteins
  JOURNAL   Patent: JP 2002523045-A 16 30-JUL-2002;
            INCYTE PHARMACEUTICALS INC
OS    Homo sapiens (human)
PN    JP 2002523045-A/16
PD    30-JUL-2002
PF    20-AUG-1999 JP 2000566425
PR    21-AUG-1998 US 60/097550,12-JAN-1999 US     60/115639 PI
JENNIFER L HILLMAN,HENRY YUE,Y TOM TANG,NEIL C CORLEY,KARL J   PI
GUEGLER,
PI    GINA A GORGONE,CHANDRA PATTERSON,MARIAH R BAUGHN,PREETI LAL,
PI    OLGA BANDMAN,
PI    ROOPA REDDY,YALDA AZIMZAI,LEO L SHIH,JUNMING
YANG,DYUNG AINA M
PI    LU
PC    C12N15/09,A61K38/00,A61K38/55,A61K45/00,A61P1/00,A61P1/04,    PC
A61P1/16,
PC    C12N15/18,A61P3/00,A61P3/10,A61P7/00,A61P7/06,A61P9/00,A61P11/
PC    00,A61P11/06,
PC    A61P13/12,A61P17/00,A61P17/06,A61P19/00,A61P19/02,A61P19/06,
PC    A61P19/10,
PC    A61P21/00,A61P21/04,A61P21/22,A61P29/00,A61P31/00,A61P31/04,
PC    A61P31/10,
PC    A61P31/12,A61P31/18,A61P31/22,A61P31/00,A61P35/02,A61P37/00,
PC    A61P37/08,
PC    C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/10 PC
,C12P21/02,C12O1/68,
PC    G01N33/15,G01N33/50,G01N33/53,G01N33/566//C12P21/08,C12N15/00,
PC    C12N5/00,
```

First alignment (continued)

```
Qy    261   ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyGly   280
Db    865   GTTCCTGGTCATTCGGGATTTATAGGCAACGAAGAGGCTGACAGATTAGCCAGAGAAGGA      924

Qy    281   AlaLysGlnSerGluAsp   286
Db    925   GCTAAACAATCGGAAGAC   942
```

RESULT 5

```
LOCUS       AF048994     1113 bp    mRNA    linear   PRI 21-NOV-1998
DEFINITION  Homo sapiens ATCC158373 ribonuclease H1 (rnh1) mRNA, complete cds.
ACCESSION   AF048994
VERSION     AF048994.1  GI:2935441
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1113)
  AUTHORS   Cerritelli,S.M., Fedoroff,O.Y., Reid,B.R. and Crouch,R.J.
  TITLE     A common forty amino acid motif in eukaryotic RNases H1 and
            caulimovirus ORF VI proteins binds to duplex RNAs
  JOURNAL   Nucleic Acids Res. (1998) In press
REFERENCE   2  (bases 1 to 1113)
  AUTHORS   Cerritelli,S.M. and Crouch,R.J.
  TITLE     Cloning, expression, and mapping of ribonucleases H of human and
            mouse related to bacterial RNase HI
  JOURNAL   Genomics 53 (3), 300-307 (1998)
  MEDLINE   99017966
   PUBMED   9799596
REFERENCE   3  (bases 1 to 1113)
  AUTHORS   Cerritelli,S.M. and Crouch,R.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1998) Laboratory of Molecular Genetics, NICHD
            NIH, 6 Center Drive Building 6B Room 2B-231, Bethesda, MD 20892,
            USA
FEATURES             Location/Qualifiers
     source          1..1113
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /strain="ATCC158373"
                     /db_xref="ATCC:158373"
                     /db_xref="dbEST:AA356084"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p11.2"
                     /cell_type="Jurkat T-cells VI"
     gene            1..1113
                     /gene="rnh1"
     CDS             28..888
                     /gene="rnh1"
                     /note="similar to bacterial RNases HI"
                     /codon_start=1
                     /product="ribonuclease H1"
                     /protein_id="AAC78563.1"
                     /db_xref="GI:2935442"
                     /translation="MSWFLFLAHRVALAALPCRRGSRGFGMFYAVRGRKTGVFLTWN
                     ECRAQVDRFPPAARFKKFATEDEAWAFVRKSASPEVSEGHENQHGQESEAKASKRLREP
                     LDGDHRSAEPVAKHMKPSVEPAPPVSRDTFSYMGDFVVVYTDGCCSSNGRRPRAGI
                     GVYWGPGHPLNVGIRLPGRQTNQRAEIHAACKAIEQAKTQNINKLVLYTDSMFTINGI
                     TNWVQGWKKNGWKTSAGKEVINKEDFVALRELTQGMDIQWMHVPGHSGFIGNEEADRL
                     AREGAKQSED"
ORIGIN
```

Alignment Scores:

Pred. No.:	3.1e-101	Length:	1113
Score:	1522.00	Matches:	281
Percent Similarity:	98.95%	Conservative:	2
Best Local Similarity:	98.25%	Mismatches:	3
Query Match:	98.45%	Indels:	0
DB:		Gaps:	0

Left column

```
PC   A61K37/02,A61K37/64
CC   A61K37 Key Identification No.: 2073417CB1
FH   source           Location/Qualifiers
FT   source           1..1150
FT                    /organism='Homo sapiens (human)'.

FEATURES             Location/Qualifiers
     source          1..1150
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
ORIGIN
```

Alignment Scores:

Pred. No.:	2.8e-101	Length:	1150
Score:	1523.00	Matches:	281
Percent Similarity:	99.30%	Conservative:	3
Best Local Similarity:	98.25%	Mismatches:	2
Query Match:	98.51%	Indels:	0
DB:		Gaps:	0

US-10-054-313-1 (1-286) x BD223780 (1-1150)

```
Qy     1   MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg      20
Db    85   ATGAGCTGGCTTCTTTCTGTTCCTTGGCCCACAGAGTCGCCTTGGCCCTTGCCCTGCCCGCC  144

Qy    21   GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgGlyThrGlyValPhe       40
Db   145   GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGCCCGAGACCGGGGTCTTT      204

Qy    41   LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys    60
Db   205   CTGACCTGGAATGAGTGCAGAGCACAGGTTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG   264

Qy    61   PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer    80
Db   265   TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA   324

Qy    81   GluGlyHisGluAsnGlnHisGlyGlnGluSerAlaLysSerProGlyLysArgLeuArg   100
Db   325   GAAGGGCATGAAGAAAATCAACATGGACAAGAATCGGAGGCGAAAGCCAGCAGGAATC     384

Qy   101   GluProLeuAspGlyHisArgSerAlaGluProValValTyrAlaLysHisMetLysPro   120
Db   385   GAGCCACTGGATGGACATAGAAGCGCAGAGCCGTATGCAAAGCACATGAAGCCG         444

Qy   121   SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal   140
Db   445   AGCATGGAGCCGGCGCCTCCAGTTAGCAGAGACACGTTTTCTACATGGGAGACTTCGTC    504

Qy   141   ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgProAlaGlyIle         160
Db   505   GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCAGCAGGAATC   564

Qy   161   GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln   180
Db   565   GGCGTTTACTGGGGCCCAGGGCACCCATCTTTAAATGTAGGCATTAGACTTCCTGGGCGCCAG 624

Qy   181   ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln   200
Db   625   ACAAACCAAAGAGCCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA  684

Qy   201   AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn   220
Db   685   AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC   744

Qy   221   TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn   240
Db   745   TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC   804

Qy   241   LysGluAspPheValAlaLeuArgGluLeuThrGlnGlyMetAspIleGlnTrpMetHis   260
Db   805   AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT   864
```

US-10-054-313-1 (1-286) x AF048994 (1-1113)

```
Qy    1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaAlaAlaLeuProCysArgArg 20
Qy   28 ATGAGCTGGTTTCTGTTCCTGGCCCACAGAGTCGGCCTTGCCGCCTGCCGCCGC 87
Db      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgLysThrGlyValPhe 40
Qy   88 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGCGCCAAGACCGGGGTCTTT 147
Db      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Qy  148 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTAAGAAG 207
Db      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy   61 PheAlaThrGluAspGluAlaTrpValArgLysSerAlaSerProGluValSer 80
Qy  208 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 267
Db      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db 268 GAAGGGCATGAAAATCAACATGGACAGAAATCGGAGGCGAAAGCCAAGCGACTCCGT 327

Qy 101 GluProLeuAsnGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db 328 GAGCCACTGAATGGAGATGGACATGAAAGCCAGCCGTATGCAAAGCACATGAAGCCG 387

Qy 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db 388 AGCGTGGAGCCGGCGCCCTCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC 447

Qy 141 ValValTyrThrAspGlyCysCysSerSerArgAsnGlyTyrArgArgLysProArgAlaGlyIle 160
Db 448 GTCGTCTACACCTGATGCGTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCGAGCAGGAATC 507

Qy 161 GlyValTyrArgTrpProGlyHisProLeuAsnValGlyIleGlyProGlyArgGln 180
Db 508 GGCGTTTACTGGGGCCAGGCCATCCTTAAATGTAGGCATTAGACTTCCTGGGGCCAG 567

Qy 181 ThrAsnGlnArgAlaGlyIleHisAlaLysLysAlaIleGluGluGlnAlaLysThrGln 200
Db 568 ACAAACCAAAGAGCCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA 627

Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Db 628 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 687

Qy 221 TrpValGlnGlyTrpLysAsnGlyThrTrpLysThrSerAlaGlyLysTrpGluValLeuAsn 240
Db 688 TGGGTTCAAGGTGGAAGAAAATGGGTGGAAGACAAGTGCCAGGGAAAGAGGTGATCAAC 747

Qy 241 LysGluAspPheValAlaAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
Db 748 AAAGAGGACTTTGTGGCAGCACTTGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 807

Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
Db 808 GTTCCTGGTCATTCGGGATTTATAGGCAATGAGGAAGCTGACAGATTAGCCAGAGAAGGA 867

Qy 281 AlaLysGlnSerGluAsp 286
Db 868 GCTAAACAATCGGAAGAC 885
```

```
RESULT 6
AF048995
LOCUS         AF048995           1147 bp    mRNA            linear      PRI 21-NOV-1998
DEFINITION    Homo sapiens ATCC159806 ribonuclease H1 (rnh1) mRNA, complete cds.
ACCESSION     AF048995
VERSION       AF048995.1  GI:2935443
KEYWORDS      .
SOURCE        Homo sapiens (human)
  ORGANISM    Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1147)
  AUTHORS   Cerritelli,S.M., Fedoroff,O.Y., Reid,B.R. and Crouch,R.J.
  TITLE     A common forty amino acid motif in eukaryotic RNases H1 and
            caulimovirus ORF VI proteins binds to duplex RNAs
  JOURNAL   Nucleic Acids Res. (1998) In press
REFERENCE   2  (bases 1 to 1147)
  AUTHORS   Cerritelli,S.M. and Crouch,R.J.
  TITLE     Cloning, expression, and mapping of ribonucleases H of human and
            mouse related to bacterial RNase HI
  JOURNAL   Genomics 53 (3), 300-307 (1998)
  MEDLINE   99017966
  PUBMED    9799596
REFERENCE   3  (bases 1 to 1147)
  AUTHORS   Cerritelli,S.M. and Crouch,R.J.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-FEB-1998) Laboratory of Molecular Genetics, NICHD
            NIH, 6 Center Drive Building 6B Room 2B-231, Bethesda, MD 20892,
            USA
FEATURES             Location/Qualifiers
     source          1..1147
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /strain="ATCC159806"
                     /db_xref="ATCC:159806"
                     /db_xref="dBEST:AA312554"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17p11.2"
                     /cell_type="Jurkat T-cells VI"
     gene            1..1147
                     /gene="rnh1"
                     62..922
     CDS             /gene="rnh1"
                     /note="similar to bacterial RNases HI"
                     /codon_start=1
                     /product="ribonuclease H1"
                     /protein_id="AAC78564.1"
                     /db_xref="GI:2935444"
                     /translation="MSWFLFLAHRVALAALPCRRGSRGFGMFYAVRRGRKTGVFLTWN
                     ECRAQVDRFPRARFKKFATEDEAWAFVRKSASPEVSEGHENOHGQSEAKASKRLREP
                     LDGDGHESAEPYAKHMKPSVEPAPPVSRDTFSYMGDFVVYTDGCCSSNGRRRPRAGI
                     GVYRWPGHPLNVGIRLEGRQTNQRAEIHAACKAIEQAKTQNINKLVLYTDSMFTNGI
                     TNWVQGWKKNGWKTSAGKEVINKEDFVALERLTQGMDIQWMHVPGHSGFIGNEEADRL
                     AREGAKQSED"
ORIGIN

Alignment Scores:
Pred. No.:                          3.3e-101           Length:          1147
Score:                              1522.00            Matches:         281
Percent Similarity:                 98.95%             Conservative:    2
Best Local Similarity:              98.25%             Mismatches:      3
Query Match:                        98.45%             Indels:          0
DB:                                                    Gaps:            9
```

US-10-054-313-1 (1-286) x AF048995 (1-1147)

```
Qy    1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaAlaAlaLeuProCysArgArg 20
Db   62 ATGAGCTGGTTTCTGTTCCTGGCCCACAGAGTCGGCCTTGCCGCCTGCCGCCGC 121
```

```
Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgLysThrGlyValPhe 40
Db  122 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGCGCCAAGACCGGGGTCTTT 181
```

```
Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db  182 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTAAGAAG 241
```

```
Qy   61 PheAlaThrGluAspGluAlaTrpValAlaAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 80
Db  242 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 301
```

```
/gene="rnhII"
/note="evolutionarily related to prokaryotic RNase HI and
RNase H domain of retroviral reverse transcriptase;
ribonuclease HII"
/codon_start=1
/product="RNase HII"
/protein_id="CAA11835.1"
/db_xref="GI:3043447"
/db_xref="GOA:O60930"
/db_xref="SWISS-PROT:O60930"
/translation="MSWLLFLAHRVALALPCRRGSRGFGMFYAVRRGRKTGVFLTWN
ECRAQVDRFPAARFKKFATEDEAWAFVRKSASPEVSEGHENQHGRESEAKASKRLREP
LDGDGHESAEPYAKHMKPSVEPAPPVSRDTFSYMGDFVVVYTDGCCSSNGRRRFRAGI
GVYWGPGHPLNVGIRLPGRQTNQRAEIHAACKAIEQAKTQNINKLVLYTDSMFTINGI
TNWVRGWKKNGWKTSAGKEVINKEDFVALERLTQGMDIQWMHVPGHSGFIGNEBADRL
AREGAKQSED"
```

ORIGIN

```
Alignment Scores:
Pred. No.:                              Length:                861
Score:              4.66e-101           Matches:               280
Percent Similarity:  99.30%             Conservative:            4
Best Local Similarity: 97.90%           Mismatches:              2
Query Match:         98.19%             Indels:                  0
DB:                                     Gaps:                    0
```

US-10-054-313-1 (1-286) x HSAJ4117 (1-861)

```
Qy      1  MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuProCysArgArg      20
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  ATGAGCTGGCTTCTGTTCCTGGCCCACCAGAGTCGTGCGCTGGCCCTTGCCCTGCCGCGC  60

Qy     21  GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe   40
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  GGCTCCCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT  120

Qy     41  LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys   60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  CTGACCTGGAATGAGTGCAGAGCACAAGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG  180

Qy     61  PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA  240

Qy     81  GluGlyHisGluAsnGlnHisGlyArgGluSerGluAlaLysAlaSerLysArgLeuArg  100
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  GAAGGGCATGAAAATCAACATGACGAGAATCGGAGCGGAAAGCGAAGCCAAGCGACTCCGT  300

Qy    101  GluProLeuAspGlyAspGlyHisGluSerAlaGluProTyrAlaLysHisMetLysPro  120
           |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db    301  GAGCCACTGGATGGAGATGGACATGGACATGAGCCCGTATGCCAAGCACATGAAGCCG  360

Qy    121  SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  AGCGTGGAGCCGGCCGCCGCCTCCAGTTAGCAGAGACACGTTTCCTACATGGAGACTTCGTC  420

Qy    141  ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgArgPheArgAlaGlyIle  160
           |||||||||||||||||||||||||||||||||:::|||||||||||||||||
Db    421  GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGGAGCAGGAATC  480

Qy    161  GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  GGGGTTTACTGGGGCCCAGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGGCAG  540

Qy    181  ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln  200
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA  600

Qy    201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  AACATCAATAAACTGGTTCTGTACACAGACAGTATGTTTACGATAAATGGTATAACTAAC  660

Qy    221  TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn  240
```

(left column, continued)

```
Qy     81  GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg  100
           ||||||||||||||||||||||
Db    302  GAAGGGCATGAAAATCAACATGACAAGAATCGGAGCGGAAAGCGAAGCCAAGCGACTCCGT  361

Qy    101  GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
           ||||||||||||||||::::|||||||||||||||||||||||||||||||
Db    362  GAGCCACTGGATGGAGATGGACATGGACATGAGCCCGTATGCCAAGCACATGAAGCCG  421

Qy    121  SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
           |||||||||||||||||||||||||||||||||||||||||||||
Db    422  AGCGTGGAGCCGGCCGCCGCCTCCAGTTAGCAGAGACACGTTTCCTACATGGAGACTTCGTC  481

Qy    141  ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgProAlaGlyIle  160
           |||||||||||||||||||||||||||||||||||::::||||||||||
Db    482  GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCGAGCAGGAATC  541

Qy    161  GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542  GGGGTTTACTGGGGCCCAGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGGCAG  601

Qy    181  ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaIleLysThrGln  200
           ||||||||||||||||||||||||||||||||||||||||||
Db    602  ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA  661

Qy    201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
           |||||||||||||||||||||||||||||||||||||
Db    662  AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC  721

Qy    221  TrpValGlnGlyTyrLysLysAsnGlyTyrTrpLysThrSerAlaGlyLysGluValIleAsn  240
           |||||||||||||||||||||||||||||||
Db    722  TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAGAGAGTGATCAAC  781

Qy    241  LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis  260
           ||||||||||||||||||||||||||||||||||||||
Db    782  AAAGAGGACTTTGTCGGCACTGGAGAGGCTTACCCCGGGGATGGACATTCAGTCGATGCAT  841

Qy    261  ValProGlyHisSerGlyPheIleGlyAsnGluAlaAspArgLeuArgGluGlyAlaGlyLy  280
           ||||||||||||||||||||||||||||||||||||||||||||||
Db    842  GTTCCTGGTCATTGGGATTTATAGGCAATGAAGAAGTGACAGATTAGCCAGAAGGA  901

Qy    281  AlaLysGlnSerGluAsp  286
           |||||||||||||||
Db    902  GCTAAACAATCGGAAGAC  919
```

```
RESULT 7
HSAJ4117
LOCUS       HSAJ4117             861 bp    mRNA    linear   PRI 07-APR-1998
DEFINITION  Homo sapiens mRNA RNase HII.
ACCESSION   AJ224117
VERSION     AJ224117.1  GI:3043446
KEYWORDS    RNase HII; rnhII gene.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 861)
  AUTHORS   Frank,P., Braunshofer-Reiter,C. and Wintersberger,U.
  TITLE     Cloning and functional expression of human RNase HII
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 861)
  AUTHORS   Frank,P., Braunshofer-Reiter,C. and Wintersberger,U.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-FEB-1998) Department of Molecular Genetics, Institute
            of Tumor Biology-Cancer Research, Borschkegasse 8a, Vienna, Austria
            A-1090
FEATURES             Location/Qualifiers
     source          1..861
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene            1..861
                     /gene="rnhII"
     CDS             1..861
```

```
Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuAlaLeuProCysArgArg  20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   30 ATGAGCTGGCTTCTTGTTCCTGGCCCACAGAGTCGCCCTTGGCCCGCCTTGCCCTGCCGCCGCGC  89

Qy   21 GlySerArgLeuPheValAlaLeuGluArgThrGlnMetPheTyrAlaValArgArgLysThrGlyValPhe  40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   90 GGCTCCTCGCGGGTTCGGGAGTTCTATGCCGTGAGGAGGGCCGCAAGAACCGGGGGTCTTT  149

Qy   41 LeuThrTrp-------  43
          ||||||||
Db  150 CTGACCTGTGAGTGGCCGC  209

Db  210 GGCCGGAGCCCGGCTAACCCCGGCGGCGGGGAGCCGGGGTCTGGCGGTCGTGCGCCGCGGC  269

Qy   43 -------  43
Db  270 TGTGGCTGGTGGTCGGGGGGACCCGCCGCTTCCTGCTCCGTTTCCTGCTCCGCGGGGGTTCCAGG  329

Qy   43 -------  43
Db  330 AGGCTCATGGCCTGGGACTGGACGCCACCGGAAGACTTGTTTGAAATATCAGAAACGTCGC  389

Qy   44 ------AsnGluCysArgAlaGlnValAlaAspArgPheProAlaAlaArgPheLysLysPh  61
Db  390 GTAAAGCGAATGAGTGCAGAGCACAGGTGACCAGGTTCCTGCTGCCAGATTTAAGAAGTT  449

Qy   61 eAlaThrGluAspGluAlaTrpAlaPheValSerAlaSerProLeuValSerGl  81
Db  450 TGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTCAAGCCCCGAAGTTTCAGA  509

Qy   81 uGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGl  101
Db  510 AGGGCATGGAAGAATCAACATGGACAAGAATCAACATGGAGGGAAGCAGCAGACTCCGTGA  569

Qy  101 uProLeuAspGlyAspProValHisGluSerAlaGlnProTyrAlaAlaLysHisMetLysProSe  121
          ||||||   | |:::
Db  570 GCCACTGGATGGAGATGGACATGAAGAGCCGTATGCAAAGCACATGAAGCCGAG  629

Qy  121 rValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheValVa  141
Db  630 CGTGGAGCCGGCCGCCTCCAGTTAGCAGAGACACGTTTCCTACATGGGAGACTTCGTCGT  689

Qy  141 lValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProAlaArgAlaGlyIleGl  161
          ||:||||
Db  690 CGTCTACACTGATGGCTGTTGCTCCAGTAATGGGCGTAGGAAGCCGGCAGCAGGAATCGG  749

Qy  161 yValTyrTrpGlyTyrHisProLeuAsnValGlyIleArgLeuProGlyArgGlnTh  181
Db  750 CGTTTACTGGGGCTACCATCCTTAAATGTAGGCATTAGGACTTCCTGGGCGGCGGCAGAC  809

Qy  181 rAsnGlnArgAlaGluIleHisAlaAlaCysLysValGluGluThrGlnAsp  201
Db  810 AAACCAAAGAGCGGAAATTCATGCAGCCTCCAAAGCCATTGAAACAAGACTCAAAA  869

Qy  201 nIleAsnLysValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTr  221
Db  870 CATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAAACTACTG  929

Qy  221 pValGlnGlyTrpLysSerAsnGlyTrpSerThrSerAlaGlyLysGluValIleAsnLy  241
Db  930 GGTTCAAGGTTGGAAGAAAATGGGTGGAGCAGGAAGAGAGTGATCAACAA  989

Qy  241 sGluAspPheValAlaLeuGluThrGlyLeuThrGlnGlyMetAspIleGlnTrpMetHisVa  261
Db  990 AGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGATGGACATTCAGTGGATGCATGT  1049

Qy  261 lProGlyHisSerGlyPheIleGlyAsnGluGlyAlaAspArgLeuAlaArgGluGluArgGlyA  281
Db 1050 TCCTGGTCATTGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAGGAGAGC  1109

Qy  281 aLysGlnSerGluAsp  286
```

```
Db  661 TGGGGTTCGAGGTTCCGAGGTTGGAAGAGAAAATGGTGAAGACAAGTGCAGGAGGAAGAGAGGAAGAGGTGATCAAC  720

Qy  241 LysGluAspPheValAlaLeuGluArgThrGlnMetPheTyrMetHis  260
Db  721 AAAGAGGACTTTGTGCACTGGAGAGGCTTACCCAGGGATGGACATTCAGTGGATGCAT  780

Qy  261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGluGly  280
Db  781 GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAGGA  840

Qy  281 AlaLysGlnSerGluAsp  286
Db  841 GCTAAACAATCGGAAGAC  858
```

RESULT 8

AK096913

LOCUS   AK096913   2120 bp   mRNA   linear   PRI 15-JUL-2002

DEFINITION   Homo sapiens cDNA FLJ39594 fis, clone SKNSH2001875, highly similar to Homo sapiens ATCC158373 ribonuclease H1 (rnh1) mRNA.

ACCESSION   AK096913

VERSION   AK096913.1   GI:21756512

KEYWORDS   oligo capping; fis (full insert sequence).

SOURCE   Homo sapiens (human)

ORGANISM   Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE   1

AUTHORS   Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R., Otsuki,T., Sato,H., Wakamatsu,A., Ishii,S., Yamamoto,J., Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K., Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H., Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K., Takahashi-Fujii,A., Oshima,A., Suzuki,Y., Sugano,S., Nagahari,K., Masuho,Y., Nagai,K. and Isogai,T.

TITLE   NEDO human cDNA sequencing project

JOURNAL   Unpublished

REFERENCE   2 (bases 1 to 2120)

AUTHORS   Isogai,T. and Yamamoto,J.

TITLE   Direct Submission

JOURNAL   Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0812, Japan (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)

COMMENT   NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan; cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- and one pass sequencing: RAB; 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

FEATURES   Location/Qualifiers
    source   1..2120
       /organism="Homo sapiens"
       /mol_type="mRNA"
       /db_xref="taxon:9606"
       /clone="SKNSH2001875"
       /cell_line="SK-N-SH"
       /cell_type="neuroblastoma"
       /clone_lib="SKNSH2"
       /note="cloning vector: pME18SFL3"

ORIGIN

Alignment Scores:			
Pred. No.:	2.98e-96	Length:	2120
Score:	1457.50	Matches:	282
Percent Similarity:	77.60%	Conservative:	2
Best Local Similarity:	77.05%	Mismatches:	2
Query Match:	94.28%	Indels:	80
DB:	9	Gaps:	1

US-10-054-313-1 (1-286) x AK096913 (1-2120)

us-10-054-313-1.p2n.rge

```
Db      ||||||||| ||||||||||
        1110 TAAACAATCGGAAGAC 1125
```

RESULT 9
AF048993

LOCUS	AF048993 1430 bp mRNA linear ROD 21-NOV-1998
DEFINITION	Mus musculus ribonuclease H1 (Rnh1) mRNA, complete cds.
ACCESSION	AF048993
VERSION	AF048993.1 GI:2935417
KEYWORDS	.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1 (bases 1 to 1430)
AUTHORS   Cerritelli,S.M., Fedoroff,O.Y., Reid,B.R. and Crouch,R.J.
TITLE   A common forty amino acid motif in eukaryotic RNases H1 and
            caulimovirus ORF VI proteins binds to duplex RNAs
JOURNAL   Nucleic Acids Res. (1998) In press
REFERENCE   2 (bases 1 to 1430)
AUTHORS   Cerritelli,S.M. and Crouch,R.J.
TITLE   Cloning, expression, and mapping of ribonucleases H of human and
            mouse related to bacterial RNase HI
JOURNAL   Genomics 53 (3), 300-307 (1998)
MEDLINE   99017966
PUBMED   9799956
REFERENCE   3 (bases 1 to 1430)
AUTHORS   Crouch,R.J. and Cerritelli,S.M.
TITLE   RNases H of lower eukaryotes: Saccharomyces cerevisiae,
            Schizosaccharomyces pombe, Neurospora crassa and Crithidia
            fasiculata
JOURNAL   (in) Crouch,R.J. and Toulme,J.-J. (Eds.);
            RIBONUCLEASES H;
            INSERM Editions, Paris, France (1998) In press
REFERENCE   4 (bases 1 to 1430)
AUTHORS   Cerritelli,S.M. and Crouch,R.J.
TITLE   Direct Submission
JOURNAL   Submitted (12-FEB-1998) Laboratory of Molecular Genetics, NICHD,
            NIH, 6 Center Drive, Building 6B, Room 2B-231, Bethesda, MD 20892,
            USA
FEATURES   Location/Qualifiers
     source   1..1430
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="dbEST:AA163387"
                     /db_xref="taxon:10090"
                     /chromosome="12"
                     /map="12A3"
                     /tissue_type="carcinoma"
                     /clone_lib="Stratagene 937317"
                     /dev_stage="embryonic"
     gene   1..1430
                     /gene="Rnh1"
     5'UTR   1..49
                     /gene="Rnh1"
     CDS   49..906
                     /gene="Rnh1"
                     /note="similar to bacterial RNases HI"
                     /codon_start=1
                     /product="ribonuclease H1"
                     /protein_id="AAC78562.1"
                     /db_xref="GI:2935418"
                     /translation="MRWLLPLSRTVTLAVVRLRRGICGLGMFYAVRRGRRTGVFLSWS
                     ECKAQVDRFPAARFKKFATEDEAWAFVRSSSSPDGSKGQESAHEQKSQAKTSKRPREP
                     LGEGEELPEPGPKHTRQDTEPAAVVSKDTFSYMGESVIVYTDGCCSSNGRKRARAGIG
                     VYWGPGHPLNVGIRLPGRQTNQRAEIHAACKAIMQAKAQNISKLVLYTDSMFTINGIT
                     NWVQGWKKNGWRTSTGKDVINKSDFMELDELTQGMDIQWMHIPGHSGFVGNEEADRLA
                     REGAKQSED"
     3'UTR   907..1430
                     /gene="Rnh1"
ORIGIN
```

Alignment Scores:

Pred. No.:	
Score:	1197.50
Percent Similarity:	86.71%
Best Local Similarity:	77.27%
Query Match:	77.46%
DB:	1.16e-77

Length:	1430
Matches:	221
Conservative:	27
Mismatches:	37
Indels:	1
Gaps:	1

US-10-054-313-1 (1-286) x AF048993 (1-1430)

```
Qy      1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg  20
          ||  || ||     || |:||   |::|| ||        ||
Db     49 ATGGGCTGGCTGCTGCCGCTGTCCCCACAGTGACACTGGCCGTGCTACGCTGAGGCGA 108

Qy     21 GlySerArgGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe  40
          ||||||  ||||  |    :|||| |::||
Db    109 GGCATTTGCCGCTCGGAGTGTTCTATGGCGGTGAGGAGGAGGAGCCGACCGGAGTCTTC 168

Qy     41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys  60
          |:||:||||  || || | || ||  ||| ||| || || |::|||:|||
Db    169 CTGAGTTGGAGTGCAAAGCCCAGGTGGACCGGTTTCCTGCTGCCAGGTTTAAGAAA 228

Qy     61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer  80
          |:|| ||  |||| || ::|| ||||  ||:|::|| ::|::
Db    229 TTTGCCACAGAAGATGAGGCCTGGGCCTTTGTCAGGAGCTCTTCAAGCCCGGATGGTTCA 288

Qy     81 GluGlyHisGluAsnGlnHisGlyIleGlyIleGluSerGluAlaLysProGlyLysArgLeuArg  100
          :::||   | ||:|:  | |  :||: |:::  |:::
Db    289 AAAGGGCAGGAAGTGCATGACAGAAGTCACAGGCGAAGACACAGCAAGCGGCCCTCGG 348

Qy    101 GluProLeuAspGlyAspGlyHisGluSerLysGlnProTyrAlaIleProMetLysMetLysPro  120
          ||:||:||           :::
Db    349 GAGCCCCTG---GGTGAAGGGGAAGAACTTCCAGCACCAGGCCAAAGCACACACGACAG 405

Qy    121 SerValGluProAlaProProValSerSerArgGlyArgArgAspThrPheSerTyrMetGlyAspPheVal  140
          ||||||||  :|:  | |::||||     ||: ||
Db    406 GACACCGAGCCCGCTGCTGTAGTGAGCAAGGACACATTTTCTTATATGGGAGAGTCAGTC 465

Qy    141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysArgAlaArgAlaGlyIle  160
          :|| ||||:||  :|::||        :::: |||
Db    466 ATTGTCTACACGGATGGCTGTTGCTCCTCCAGTAATGGACGGGGACAGCAGAGAATT 525

Qy    161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln  180
          || |||    |||||          || |:
Db    526 GGGGTTTACTGGGGCCCAGGCCACCCCTTAAATGTAGTATAAGGCTTCCTGGGCGACAG 585

Qy    181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln  200
          || |  |||  |::|||  |:||||||:|:||
Db    586 ACAAACCAGAGGGCCGAGATCCAGCCTGCAAGGCCATCATGCAAGCCAAGGCTCAG 645

Qy    201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
          ||  ||:| ||   |||:|:|||:|| || ::|  ||  ||::
Db    646 AACATCAGCAAGCTGGTTCTGTACACAGACAGCATGTTCACCATCAATGGAGATAAACTAAC 705

Qy    221 TrpValGlnGlyTrpLysLysAsnGlyTrpArgThrSerAlaGlyLysLysGluValIleAsn  240
          ||| |  |||||        |::|::|||:::::
Db    706 TGGGTTCAGGGCTGGAAGAAGAATGGCTGGAGAACAAGTACAGGAAAGATGTGATCAAC 765

Qy    241 LysGluAspPheValAlaLeuGluLeuThrGlnGlyMetAspIleTrpMetHis  260
          ||  :::       ::::
Db    766 AAGAGGACTTCATGGAGCTGGACGAGCTCACTCAGGGCATGGACATCCAGTGGATGCAC 825

Qy    261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyGly  280
          :::   || ::|||      |||  |:|:|
Db    826 ATTCCTGGTCACTCAGGATTTGTGGGCAATGAAGAGGCCGACAGACTGGCACGGGAAGGA 885

Qy    281 AlaLysGlnSerGluAsp  286
          || :::
Db    886 GCGAAGCAGTCTGAGGAC 903
```

RESULT 10
BC019411

LOCUS	BC019411 1409 bp mRNA linear ROD 06-OCT-2003
DEFINITION	Mus musculus ribonuclease H1, mRNA (cDNA clone MGC:30220

```
ACCESSION   IMAGE:5133942), complete cds.
VERSION     BC019411
VERSION     BC019411.1  GI:18044435
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1409)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
  MEDLINE   22338257
   PUBMED   12477932
REFERENCE   2  (bases 1 to 1409)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-DEC-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Baylor College of Medicine Human Genome
            Sequencing Center
            Center code: BCM-HGSC
            Web site: http://www.hgsc.bcm.tmc.edu/cdna/
            Contact: amg@bcm.tmc.edu
            Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Loulseged, H.,
            Kowis, C.R., Sneed, A.J., Martin, R.G., Muzny, D.M., Nanavati,
            A.N., Gibbs, R.A.
  COMMENT   Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 41 Row: n Column: 3
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 6755333.
FEATURES             Location/Qualifiers
     source          1..1409
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="MGC:30220 IMAGE:5133942"
                     /tissue_type="Liver, normal, 5 month old male mouse."
                     /clone_lib="NCI_CGAP_Li9"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
     gene            1..1409
                     /gene="Rnaseh1"
                     /db_xref="LocusID:19819"
```

```
                     /db_xref="MGI:1335073"
     CDS             17..874
                     /codon_start=1
                     /product="ribonuclease H1"
                     /protein_id="AAH19411.1"
                     /db_xref="GI:18044436"
                     /db_xref="LocusID:19819"
                     /translation="MRWLLPLSRTVTLAVVLRLRRGICGLGMFYAVRRGRRTGVFLSWS
                     ECKAQVDRFPAARFKFATEDEAWAFVRSSSPPDGSKGQESAHEQKSQVKTSKRRPEP
                     LGEGEELPEPGPKHTRQDTEPAAVVSKDTFSYMGESVIVTDGCCSSNGRKRARAGIG
                     VYWGPGHPLNVGIRLPGRQTNQRAEIHAACKAIMQAKAQNISKLVLYTDSMFTINGIT
                     NWVQGWKKNGWRTSTGKDVINKEDFMELDELTQGMDIQWNHIPGHSGFVGNEEADRIA
                     REGAKQSED"
     misc_feature    89..232
                     /note="COG3341; Region: Predicted double-stranded
                     RNA/RNA-DNA hybrid binding protein [General function
                     prediction only]"
                     /db_xref="CDD:COG3341"
     misc_feature    428..859
                     /note="RnhA; Region: Ribonuclease HI [DNA replication,
                     recombination, and repair]"
                     /db_xref="CDD:COG0328"
ORIGIN

Alignment Scores:
Pred. No.:                                  Length:            1409
Score:                       2.22e-77       Matches:           220
Percent Similarity:          1193.50        Conservative:      27
Best Local Similarity:       86.36%         Mismatches:        38
Query Match:                 76.92%         Indels:            1
DB:                          77.20%         Gaps:              1
                             10
```

```
US-10-054-313-1 (1-286)  x BC019411 (1-1409)

Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaAlaLeuArgLeuArgArgGlyIleCysArgArg 20
Db   17 ATGGGCTGGCTGCTGCCGCTGTCCCGCACAGTGACACTGGCCGTCGTACGCCTGAGGCGGCGA 76

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
Db   77 GGCATTTGGCGCTCGGGATGTTCTATGCGGTGAGGAGGAGGCCCAGGAGCCGGAGTCTTC 136

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
Db  137 CTGAGTTGGAGTGAGTGCAAAGCCCAGGTGGACCCGGTTTCCTGCTGCCAGGTTTAAGAAA 196

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
Db  197 TTTGCCACAGAAGATGAGGCCTGGGCCTTTGTCAGGAGTCTTCAAGCCCGGATGGTTCA 256

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
Db  257 AAAGGGCAGGAAAGTGCACATGAGCAGAGTCACAGGTGAAGACCAGCAGGCGGCCTCGG 316

Qy  101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db  317 GAGCCCCTG---GGTGAAGGGGAAGAACTTCCAGACGCCAGGGCCAAAGCACACACGACAG 373

Qy  121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
Db  374 GACACCGAGCCCGCTCGCTGCTAGTGAGCAAGGACACGTTTTCTTATATGGGAGAGTCAGTC 433

Qy  141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgGlyLysProArgAlaGlyIle 160
Db  434 ATTGTCTACACGGATGGCTGTTGCTCCAGTAATGGACGGGGCCACGAGCCAGGAATT 493

Qy  161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
Db  494 GGGCGTTTTACTTGGGGCCCCCAGGCCACCCCCTTAAATGTAGGTATAAGGCTTCCTGGGGGACAG 553

Qy  181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysAlaGlnAsnThrGln 200
Db  554 ACAAACCAGAGGCCGAGATCCATGCAGCTGCCAAGGCCATCATGCAGGCCAAGGCTCAG 613
```

```
Qy   201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
          |||||:|||||:|||::||||||||||||||||||||||||||||||||
Db   614  AACATCAGCAAGTTGGTTCTGTACACAGACAGCAGCATGTTCACCATCAATGGGATAACTAAC 673

Qy   221  TrpValGlnGlyIleTrpLysLysAsnGlyTyrLysTrpLysThrSerAlaGlyLysGluValIleAsn 240
          ||||||||||||:|||::|||||:||||||||||||||
Db   674  TGGGTTCAGGGCTGGAAGAAGAATGGCTGGAGAACAACGTACAGGGAAGATGTGATCAAC 733

Qy   241  LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis 260
          |||::|||::||
Db   734  AAGGAGGACTTCATGGAGCTGGACGAGCTCACTCAGGGCATGGACATCCAATGGATGCAC 793

Qy   261  ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGlyGly 280
          ::||||||||||:|||
Db   794  ATTCCTGGTCACTCAGGATTTGTGGGCAATGAAGAGGCCGACAGACTGGCACGGGAAGGA 853

Qy   281  AlaLysGlnSerGluAsp 286
          ||||||||||||||
Db   854  GCGGAGCAGTCTGAGGAC 871
```

```
RESULT 11
AC098850
LOCUS       AC098850               178965 bp    DNA     linear   PRI 30-DEC-2001
DEFINITION  Homo sapiens chromosome 17, clone RP11-92B11, complete sequence.
ACCESSION   AC098850
VERSION     AC098850.2  GI:17646937
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 178965)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-92B11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178965)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., Pierre,N.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 178965)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., Pierre,N.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Dec 13, 2001 this sequence version replaced gi:16604041.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------------  Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submission@genome.wi.mit.edu
            ------------------  Project Information
            Center project name: L13829
            Center clone name: 92_B_11
            ------------------
FEATURES             Location/Qualifiers
     source          1..178965
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-92B11"
                     /clone_lib="RPCI-11 Human Male BAC"
     unsure          987..1002
                     /note="<30 qual SNGL region"
     repeat_region   994..1308
                     /rpt_family="AluJo"
     repeat_region   complement(1434..1735)
                     /rpt_family="AluSx"
     repeat_region   2075..2095
                     /rpt_family="AT_rich"
     repeat_region   complement(2205..2509)
                     /rpt_family="AluSx"
     repeat_region   2558..2592
                     /rpt_family="AluSx"
     repeat_region   2730..3033
                     /rpt_family="AT_rich"
     repeat_region   3034..3339
                     /rpt_family="AluJo"
     repeat_region   4237..4520
                     /rpt_family="AluSx"
     repeat_region   4584..5144
                     /rpt_family="AluSx"
     repeat_region   5180..5481
                     /rpt_family="LIME"
     repeat_region   complement(5496..5542)
                     /rpt_family="AluSx"
     repeat_region   complement(5543..5678)
                     /rpt_family="AluSx"
     repeat_region   complement(5801..6120)
                     /rpt_family="AluSx"
     repeat_region   6123..6363
                     /rpt_family="MER58B"
     repeat_region   6423..6533
                     /rpt_family="LIME"
     repeat_region   6534..6830
                     /rpt_family="AluSx"
     repeat_region   6831..7175
                     /rpt_family="LIME"
```

```
repeat_region    7310..7435        /rpt_family="L1ME"
repeat_region    7527..7703        /rpt_family="L1ME"
repeat_region    7704..7993        /rpt_family="AluJo"
repeat_region    7994..8309        /rpt_family="L1ME"
repeat_region    8319..8807        
repeat_region    8893..9579        /rpt_family="L2"
repeat_region    9941..9994        /rpt_family="L2"
repeat_region    complement(10332..10491)  /rpt_family="GA-rich"
repeat_region    complement(11158..11392)  /rpt_family="MER3"
repeat_region    11639..12721      /rpt_family="L3"
repeat_region    12814..13108      /rpt_family="L1MC5"
repeat_region    13131..13216      /rpt_family="AluY"
repeat_region    complement(13203..13257)  /rpt_family="L1P"
repeat_region    13255..15125      /rpt_family="L1PA4"
repeat_region    complement(15190..15383)  /rpt_family="L1PA4"
repeat_region    15752..16394      /rpt_family="MIR"
repeat_region    16392..16515      /rpt_family="L1MB7"
repeat_region    16513..17530      /rpt_family="L1MB7"
repeat_region    complement(17531..17840)  /rpt_family="AluSx"
repeat_region    17841..19946      /rpt_family="AluSg"
repeat_region    20113..20529      /rpt_family="L1MB7"
repeat_region    20934..21416      /rpt_family="MSTB1"
repeat_region    21417..21594      /rpt_family="LTR65"
repeat_region    21929..22229      /rpt_family="AluSg"
repeat_region    complement(22689..23156)  /rpt_family="AluSx"
repeat_region    complement(24183..24502)  /rpt_family="LTR37B"
repeat_region    complement(26205..26386)  /rpt_family="AluSx"
repeat_region    26671..26697      /rpt_family="MIR"
repeat_region    27404..27705      /rpt_family="(TAA)n"
repeat_region    complement(28018..28146)  /rpt_family="AluSx"
repeat_region    28157..28214      /rpt_family="MER5A"
repeat_region    28714..28743      /rpt_family="AT_rich"
repeat_region    28825..28870      /rpt_family="AT_rich"
repeat_region    31073..31292      /rpt_family="MIR"
repeat_region    31553..31858      /rpt_family="L1ME4A"
repeat_region    32320..32351      /rpt_family="L3"
repeat_region    32720..33020      /rpt_family="AT_rich"
```

```
repeat_region    /rpt_family="AluJo"
                 33053..33370
repeat_region    /rpt_family="AluYb8"
                 complement(33693..34052)

Alignment Scores:
Pred. No.:                            Length:          178965
Score:                   1.24e-73     Matches:         236
Percent Similarity:      1174.00      Conservative:    7
Best Local Similarity:   88.36%       Mismatches:      18
Query Match:             85.82%       Indels:          16
DB:                      75.94%       Gaps:            3
                         9

US-10-054-313-1 (1-286) x AC098850 (1-178965)

Qy     1 MetSerTrpLeuLeuPheLeuLeuAlaHisArgValAlaAlaAlaLeuLeuProCysArgArg 20
         | || |  ||| |  |   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 59696 ATGAGCCGGCTTCTGTCCCTGGCGGC-AGAGTCGCCTTGGCCCGCCCTTGCCCTGCCGCCGC 59754

Qy    21 GlySerArgGlyPheGlyMetPheTyrAlaValArgAlaArgLysThrGlyLysValPhe 40
         ||| |  |  |  |  | |  |  |  |  |  |  | ::: |  |  |  |  |  |
Db 59755 --TCTCGCGGGTTCGGGATGTTCTATCCGTGAGGAGGGCCGCAAGTCCGGGGTCTTT 59811

Qy    41 LeuThrTrpAsnGluCysArgAlaGlnValAlaAspArgPheProAlaAlaArgPheLysLys 60
         ||| |  |  |  |  | ||:|| |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 59812 CTGACCGGGAATGAGTGCAAAGCACAGGTGGACCGGTTTCCTCGTCGCCAGATTTAAGAAG 59871

Qy    61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerLaSerProGluValSer 80
         ||| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 59872 TTTGCCACAGAGGATGAGGCCTGGACTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 59931

Qy    81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysProGlyLeuArg 100
         ||| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 59932 GAAGGGCAGGAGAAAATCAACATGACAAGAATCGGAGACGAAGCCAGCAAGCGACTCCGT 59991

Qy   101 GluProLeuAspAspGlyHisGluSerAlaGlnProTyrAlaLysHisMet------ 118
         ||| |  |  |  |  |  |  |  |  |  |  |  |  ::::|  |  |
Db 59992 GAGCCACTGGATGCGATGGAGATCGAGATGAAAGCGCAGAGCCATATGC-AAGCACATGAAGCCG 60050

Qy   119 ------LysProSerValGluProAlaProProValSerArgArgAspThr 132
               |  |  | :::| |  |  |  |  |  |  |  |  |  |  |  |
Db 60051 AGCGTGAAGCCGAGCCGTGAAGCCGTGAAGCGTGAAGCCTCCAGTAGCAGAGGACACCG 60110

Qy   133 PheSerTyrMetGlyAspPheValValTyrThrAspGlyCysCysSerSerAsnGly 152
         ||| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 60111 TTTTCCTACATGGGAGACTTCGTCGTCTACCGTGATGGCTGCTGCTCCAGTAATGGG 60170

Qy   153 ArgArgLysProArgAlaGlyIleGlyIleGlyValTyrTrpGlyProGlyHisProLeuAsnVal 172
         ||| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | ||
Db 60171 CGTAGAAAGCCACGAGCAGGAATCCGCGTTTACTGGGGCCCGGGCTATCCTTTAACTGTA 60230

Qy   173 GlyIleArgLeuProGlyArgGlnThrArgGlnAlaArgAlaGlyIleHisAlaAlaCysLeuLys 192
         ||| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 60231 GGCATTAGACTTCCTGGGTGGCAGACAAACCAAAGAGCGGAAATTCATGCAGCCTGCGAAG 60290

Qy   193 AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet 212
         ||| |  |  |  |  |  |  |  |  |:::|: |  |  |  |  |  |  |  |  |
Db 60291 GCCATTGAACAAGCACAGACTCAGAAAATCAATAAACTGGTTCTGTATACAGACAGTATG 60350

Qy   213 PheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpTrpLysAsnGlyTyrTrpLysThr 232
         ||| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 60351 TTTACGATAAATGGTATAACTAACTGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACA 60410

Qy   233 SerAlaGlyLysGluValIleValAspAspProPheValAlaLeuGluGlyArgLeuThrGln 252
         ||| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 60411 AGTGCAGGAAAGGAAGAGGTGATCAACAAAGAGGACTTCGTGGCACTGGAGGAGGCTGCACCCAAG 60470

Qy   253 GlyMetAspIleGlnTrp----------MetHisValPro 262
         ||| |  |  |  |  | :::| |  |  |  |  |  |
Db 60471 GGGGATGGACATTCAGTGGCGGCGAGTATCTTAAAATGTTCATGTCCCA 60515

RESULT 12
AC022596/c
```

```
LOCUS        AC022596         192337 bp    DNA       linear    PRI 23-SEP-2000
DEFINITION   Homo sapiens chromosome 17, clone RP11-219A15, complete sequence.
ACCESSION    AC022596
VERSION      AC022596.9  GI:10280853
KEYWORDS     HTG.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 192337)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE      Homo sapiens chromosome 17, clone RP11-219A15
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 192337)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
             Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
             Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
             Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
             DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
             Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
             Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
             Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
             Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
             Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
             McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
             Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
             Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
             Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
             Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
             Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
             Zimmer,A. and Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE    3  (bases 1 to 192337)
  AUTHORS    Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
             Anderson,S., Barna,N., Bastien,V., Beda,F., Boguslavkiy,L.,
             Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
             Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
             DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,
             FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,
             Graham,L., Grand-Pierre,N., Hagos,B., Heaford,A., Horton,L.,
             Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRocque,K.,
             Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G.,
             Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,
             McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
             Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,
             O'Donnell,P., O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K.,
             Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,
             Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severy,P.,
             Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
             Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
             Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
             Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
             Zimmer,A. and Zody,M.
  TITLE      Direct Submission
  JOURNAL    Submitted (23-SEP-2000) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On Sep 23, 2000 this sequence version replaced gi:9795639.
             All repeats were identified using RepeatMasker.
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             -------------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
             --------------  Project Information
               Center project name: L4191
               Center clone name: 219_A_15
             --------------
FEATURES             Location/Qualifiers
     source          1..192337
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-219A15"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   568..788
                     /rpt_family="MLT1C"
     repeat_region   1804..1957
                     /rpt_family="FRAM"
     repeat_region   1964..2014
                     /rpt_family="MIR"
     repeat_region   complement(2029..2319)
                     /rpt_family="LTR16C"
     repeat_region   2463..2535
                     /rpt_family="MIR3"
     repeat_region   complement(2614..2700)
                     /rpt_family="MIR"
     repeat_region   complement(2938..3185)
                     /rpt_family="L1MC/D"
     repeat_region   complement(4802..4869)
                     /rpt_family="L2"
     repeat_region   complement(4871..5058)
                     /rpt_family="LTR16B"
     repeat_region   complement(5150..5227)
                     /rpt_family="LTR16A"
     repeat_region   5353..5597
                     /rpt_family="AluJo"
     repeat_region   5697..5863
                     /rpt_family="MLT1J"
     repeat_region   complement(5909..6210)
                     /rpt_family="AluSx"
     repeat_region   complement(6235..6609)
                     /rpt_family="L1MB3"
     repeat_region   6610..6687
                     /rpt_family="AluSg/x"
     repeat_region   complement(6688..6802)
                     /rpt_family="L1MB3"
     repeat_region   7399..7751
                     /rpt_family="MIR"
     repeat_region   7977..8102
                     /rpt_family="MSTD"
     repeat_region   complement(8265..8596)
                     /rpt_family="MER5A"
     repeat_region   complement(10173..10360)
                     /rpt_family="L1MB3"
     repeat_region   complement(10370..10452)
                     /rpt_family="MIR"
     repeat_region   10648..11010
                     /rpt_family="L1ME4A"
     repeat_region   complement(11168..11273)
                     /rpt_family="L1MC4"
     repeat_region   complement(11274..11583)
                     /rpt_family="AluSx"
     repeat_region   complement(11584..11603)
                     /rpt_family="L1MC4"
     repeat_region   12176..12289
                     /rpt_family="MIR"
     repeat_region   complement(12402..12464)
                     /rpt_family="L2"
     repeat_region   13177..13351
                     /rpt_family="MER69"
     repeat_region   13950..14083
                     /rpt_family="MIR"
     repeat_region   complement(14660..15282)
                     /rpt_family="L1MC4a"
     repeat_region   complement(15834..15968)
                     /rpt_family="MIR3"
     repeat_region   complement(16176..16237)
                     /rpt_family="L2"
     repeat_region   17027..17098
                     /rpt_family="MIR"
```

```
repeat_region   complement(17196..17328)
                /rpt_family="L2"
repeat_region   complement(17390..17492)
                /rpt_family="L2"
repeat_region   18087..18429
                /rpt_family="L2"
repeat_region   18524..18675
                /rpt_family="MER3"
repeat_region   18848..19158
                /rpt_family="AluSp"
repeat_region   complement(19554..19800)
                /rpt_family="AluSx"
repeat_region   complement(20989..21284)
                /rpt_family="AluY"
repeat_region   21286..21365
                /rpt_family="MER3"
repeat_region   21395..21576
                /rpt_family="L2"
repeat_region   21627..21701
                /rpt_family="MIR"
repeat_region   21770..22001
                /rpt_family="LTR14A"
repeat_region   complement(22014..22401)
                /rpt_family="L1MA4A"
repeat_region   23317..23339
                /rpt_family="AT_rich"
repeat_region   complement(23367..23502)
                /rpt_family="MIR"
repeat_region   24118..24141
                /rpt_family="(TCTG)n"
repeat_region   complement(24473..24616)
                /rpt_family="MIR3"
repeat_region   complement(25652..26053)
                /rpt_family="L2"
repeat_region   28293..28376
                /rpt_family="MIR3"
repeat_region   34302..34408
                /rpt_family="MIR"
repeat_region   complement(35693..35781)
                /rpt_family="MIR"
repeat_region   37220..37405
                /rpt_family="L2"
repeat_region   complement(37431..37554)
                /rpt_family="MIR"
repeat_region   38931..39908
                /rpt_family="LTR5"
repeat_region   40343..40448
                /rpt_family="MIR"
repeat_region   40552..40629
                /rpt_family="MIR"
repeat_region   complement(41149..41275)
                /rpt_family="MIR"
repeat_region   complement(41355..41506)
                /rpt_family="MIR3"

Alignment Scores:
Pred. No.:
Score:                  1174.00
Percent Similarity:     88.36%
Best Local Similarity:  85.82%
Query Match:            75.94%
DB:

Length:        192337
Matches:       236
Conservative:  7
Mismatches:    18
Indels:        16
Gaps:          3

1.34e-73

US-10-054-313-1 (1-286) x AC022596 (1-192337)

Qy      1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 191561 ATCAGCCCGACCTTCTGTCCCTGGCGGC-AGAGTCGGCCCTTGGCCCTGCCCGCCGCGGC 191503

Qy     21 GlySerArgGlyMetPheTyrAlaValArgGlyArgGlyThrGlyValPhe 40
            |||||||||||||||||||||||||||||||||||||::|||  ||||||||
Db 191502 ---TCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCGCAAGTCCGGTCTTT 191446
```

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAlaAspArgPheProAlaAlaArgPheLysLys 60
            ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db 191445 CTGACCCGGAATGAGTGCAAAGCACAGGTGGACCCGGTTCCTGCTGCCAGATTTAAGAAG 191386

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 191385 TTTGCCACAGAGGATGAGGCCTGGGCACTTTGTCAGGAAGTCTGCAAGCCCGGAGTTTCA 191326

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGlnAlaLysProGlyLysArgLeuArg 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 191325 GAAGGGCAGGAAATCAACATGGACAAGAATCGGAGACAGAAAGCCAGCAAGCGACTCCGT 191266

Qy   101 GluProLeuAspGlyHisGluSerAlaGlnProTyrAlaLysHisMet----- 118
            |||||||||||||||||||||||||||||:|||
Db 191265 GAGCCACTGGATGGACATGAGATGAGAAGGCCAGAGCCCATATGC-AAGCACATGAAGCCG 191207

Qy   119 ------------LysProSerValGlyProAlaProProValSerArgArgAspThr 132
                        |||||||||||||:|||
Db 191206 AGCGGTGAAGCCAGCGTGAAGCCGCGTGAAGCCGCCTCCAGTTAGCAGAGACACG 191147

Qy   133 PheSerTyrMetGlyAspPheValValValTyrThrArgGlyCysCysSerSerAsnGly 152
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 191146 TTTTCCTACATGGAGACTTCGTCGTCGTCTACGCTGATGGCTGCTGCTCCAGTAATGGG 191087

Qy   153 ArgArgLysProAlaArgAlaGlyIleGlyValTyrThrProGlyHisProLeuAsnVal 172
            |||||||||||||||||||||||||||||||||||||||||||||:::|||
Db 191086 CGTAGAAAGCCACAGAGCAGGAATCCGGTTTACTGGGGCCGGGTATCCTTTAACTGTA 191027

Qy   173 GlyIleArgLeuProGlyArgGlnThrArgAsnGlnAsnGlnAlaGluIleHisAlaAlaCysLys 192
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 191026 GGCATTAGACTTCCTGGGTGGCAGACAAACCAAAGAGCCGAAATTCATGCAGCCTGCAAG 190967

Qy   193 AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuTyrThrAspSerMet 212
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 190966 GCCATTGAACAAGCACAGACTCAGAAAATCAATAAACTGGTTCTGTATACAGAGCAGTATG 190907

Qy   213 PheThrIleAsnGlyIleThrAsnTrpValGlnGlyTyrValLysAsnGlyTrpLysThr 232
            ||||||||||||||||||||||||||||||||:||||
Db 190906 TTTACGATAAATGGTATAACTAACTGGGTTCAAGGTTGGAAGAAAATGGTGGGAAGACA 190847

Qy   233 SerAlaGlyValIleLeuAsnLysGluAspPheValAlaLeuGluArgLeuThrGln 252
            ||||||||||||||||||||||||||||||||||||||||
Db 190846 AGTGCAGGAAAGAGGTGATCAACAAAGAGGACTTCGTGGCACTTGGAGAGGCTGACCCAG 190787

Qy   253 GlyMetAspIleGlnTrp----------------MetHisValPro 262
            |||||||||||||||| :::|||||
Db 190786 GGGATGGACATTCAGTGGGCGGAGTATCTTAAATGTTCAGTCCCA 190742

```
RESULT 13
AC126352/c
LOCUS       AC126352              211403 bp    DNA     linear   PRI 19-MAR-2003
DEFINITION  Homo sapiens chromosome 17, clone CTD-2509K7, complete sequence.
ACCESSION   AC126352
VERSION     AC126352.6  GI:2912421
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 211403)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTD-2509K7
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 211403)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Devar,K., Diaz,J., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Grahan,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
```

```
  AUTHORS   McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 211403)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D. Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson.R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol.R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 211403)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., MacLean,C.,
            Lindblad-Toh,K., Liu,G., Lui,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Major,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol.R., Norbu,C., O'Connor,T., O'Donnell,P., Pierre,N.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Wu,X.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-MAR-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
  COMMENT   On Mar 19, 2003 this sequence version replaced gi:25140067.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
            Center project name: L27684
            Center clone name: 2509_K_7

FEATURES             Location/Qualifiers
     source          1..211403
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-2509K7"
                     /clone_lib="CITD2 Human BAC"
     repeat_region   3..67
                     /rpt_family="MER5A"
     repeat_region   complement(230..561)
                     /rpt_family="LIME3"
     repeat_region   1679..1776
                     /rpt_family="L3b"
     repeat_region   complement(2138..2325)
                     /rpt_family="MIR"
     repeat_region   complement(2335..2417)
                     /rpt_family="MIR"
     repeat_region   2613..2966
                     /rpt_family="LIME4a"
     repeat_region   complement(3133..3237)
                     /rpt_family="LIMC4"
     repeat_region   complement(3239..3547)
                     /rpt_family="AluSx"
     repeat_region   4140..4268
                     /rpt_family="MIR"
     repeat_region   complement(4365..4427)
                     /rpt_family="L2"
     repeat_region   5140..5314
                     /rpt_family="MER69"
     repeat_region   5913..6046
                     /rpt_family="MIR"
     repeat_region   complement(6623..7245)
                     /rpt_family="LIMC4a"
     repeat_region   complement(7797..7931)
                     /rpt_family="MIR3"
     repeat_region   complement(8139..8200)
                     /rpt_family="L2"
     repeat_region   8990..9061
                     /rpt_family="MIR"
     repeat_region   complement(9159..9291)
                     /rpt_family="L2"
     repeat_region   complement(9353..9455)
                     /rpt_family="L2"
     repeat_region   complement(9570..9638)
                     /rpt_family="L3b"
     repeat_region   10050..10392
                     /rpt_family="L2"
     repeat_region   10487..10638
                     /rpt_family="MER3"
     repeat_region   10811..11121
                     /rpt_family="AluSp"
     repeat_region   complement(11517..11763)
                     /rpt_family="AluSx"
     unsure          11534..11544
                     /note="<30 qual SNGL region"
     unsure          11554..11592
                     /note="<30 qual SNGL region"
     repeat_region   complement(12952..13247)
                     /rpt_family="AluY"
     repeat_region   13249..13328
                     /rpt_family="MER3"
     repeat_region   13356..13539
                     /rpt_family="L2"
     repeat_region   13590..13664
                     /rpt_family="MIR"
     repeat_region   13733..13964
                     /rpt_family="LTR14A"
     repeat_region   complement(13965..14364)
                     /rpt_family="L1PB4"
     repeat_region   15280..15302
                     /rpt_family="AT_rich"
     repeat_region   complement(15330..15465)
                     /rpt_family="MIR"
```

```
repeat_region    16082..16105
                 /rpt_family="(TCTG)n"
repeat_region    complement(16437..16580)
                 /rpt_family="MIR3"
repeat_region    complement(17671..18017)
                 /rpt_family="L2"
repeat_region    complement(18266..18369)
                 /rpt_family="L3b"
repeat_region    20257..20340
                 /rpt_family="MIR"
repeat_region    26262..26368
                 /rpt_family="MIR"
repeat_region    complement(27651..27739)
                 /rpt_family="MIR"
repeat_region    29178..29363
                 /rpt_family="L2"
repeat_region    complement(29389..29512)
                 /rpt_family="MIR"
repeat_region    30888..31865
                 /rpt_family="LTR5_Hs"
repeat_region    32286..32405
                 /rpt_family="MIR"
repeat_region    32509..32586
                 /rpt_family="MIR"
repeat_region    complement(33106..33232)
                 /rpt_family="MIR"
repeat_region    complement(33312..33463)
                 /rpt_family="MIR3"
repeat_region    34251..34325
                 /rpt_family="(TGG)n"
repeat_region    complement(35953..36046)
```

```
Alignment Scores:
Pred. No.:
Score:                   6.66e-73          Length:         211403
                         1165.00           Matches:        235
Percent Similarity:      88.00%            Conservative:   7
Best Local Similarity:   85.45%            Mismatches:     19
Query Match:             75.36%            Indels:         16
DB:                                        Gaps:           3

US-10-054-313-1 (1-286) x AC126352 (1-211403)

Qy     1 MetSerTrpLeuPheLeuAlaHisArgValAlaAlaLeuAlaAlaLeuProCysArgArg 20
           ||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db 183462 ATGAGCCGGCTTCGTCCCTGGCGGC-AGAGTCGGCGCCTTGGCCGCCTTGCCCTGCCGCCGC 183404

Qy    21 GlySerArgArgGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
           |||||||||||||||   ||||||||||||||||   |||
Db 183403 ---TCTCGCGGGGTCGGGATGTTCTATGCCCGTGAGGAGGGGCCAAGTCCGCCGGTCTTT 183347

Qy    41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
           |||| |||||||||||||:::|||||||
Db 183346 CTGACCGGGAATGAGTGCAAAGCACAGGTGGACCGGTTCCTGCTGCCAGATTTAAGAAAG 183287

Qy    61 PheAlaThrGluAspPheAlaPheValArgLysSerAlaSerProGluValSer 80
           |||||||||||||||||||||||||||||||||||||||||||
Db 183286 TTTGCCACAGAGGATGAGGCCTGGAGACTTTGTCAGGAAATCTGCAAGCCCCGGAAGTTCA 183227

Qy    81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
           ||||||||||:::|||||||||||||||:||||||||||||||||||
Db 183226 GAAGGGCAGGAAAATCAACATGACAAGAATCGGAGACGAAAGCCAGCAAGGACTCCGT 183167

Qy   101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMet------ 118
           |||||||||||||||:::|||||||||:::
Db 183166 GAGCCACTGGATGGAGATGGAGATGGAGCGCAGAGCCCATATGC-AAGCACATGAAGCCG 183108

Qy   119 -------LysProSerValGluProAlaProProValSerArgArgAspThr 132
           ||||||||:::||||||||||||
Db 183107 AGTGTGAAGCCAGCGGTGAAGCCGAGCGTGAAGCCGCGCCTCCCAGTTAGCAGAGACACG 183048

Qy   133 PheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGly 152
           |||||||||||||||||||||||||||||||
Db 183047 TTTTCCACATGGGAGACTTCGTCGTCGTCTACGCTGATGGCTGCTGCTCCTCCAGTAATGGG 182988

Qy   153 ArgArgGlyProArgAlaGlyLyIleGlyValAlaTyrTrpGlyHisProLeuAsnVal 172
           ||||||||||||||||||||||||||||:::||||||
Db 182987 CGTAGAAAGCCACGAGCAGCAGGAATCCGCGTTACTGGGGCGGGCTATCCTTAACTGTA 182928

Qy   173 GlyIleArgLeuProGlyArgGlyThrArgGlnArgAlaGluIleHisAlaAlaCysLys 192
           |||||||||||||||||||||||||||||||||||||||||||||||
Db 182927 GGCATTAGACTTCCTGGGTGCAGACAATCCAAAGAGCCGAAATCATCAGCCTGCAAG 182868

Qy   193 AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysValLeuTyrThrAspSerMet 212
           |||||||:::||||||||||||||||||
Db 182867 GCCATTGACCAAGCACAGACTCAGAAAATCAATAAACTGGTTCTGTATACAGACAGTATG 182808

Qy   213 PheThrIleAsnGlyIleThrAsnIleValGlnGlyTrpLysLysAsnGlyTyrPheLysThr 232
           ||||||||||||||||||||||||||||||||||
Db 182807 TTTACGATAAATGGTATAACTAACTGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACA 182748

Qy   233 SerAlaGlyLysGluValIleGluAsnLysGluAspPheValAlaLeuGluLeuThrGln 252
           |||||||||||||||||||||||||||||||||||
Db 182747 AGTGCAGGAAAGAGAGTGATCAACAAGAGGACTTCGTGCCACTGGAGAGGCTGACCCAG 182688

Qy   253 GlyMetAspIleGlnTrp------------MetHisValPro 262
                                         :::
Db 182687 GGGATGGACATTCAGTGGGGGAGTATCTTAAATGTTCATGTCCCA 182643
```

```
RESULT 14
AC090774
LOCUS       AC090774        207418 bp    DNA    linear   PRI 28-JUN-2002
DEFINITION  Homo sapiens chromosome 17, clone RP11-344E13, complete sequence.
ACCESSION   AC090774
VERSION     AC090774.6  GI:21617728
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 207418)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone RP11-344E13
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 207418)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B., Brown,A.,
            Camarata,J., Campopiano,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J.,
            Gardyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,R.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Karatas,A., LaRocque,K., Lamazares,R., Landers,T.,
            Lehoczky,J., Levine,R., Liu,G., MacLean,C., Macdonald,P.,
            Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K.,
            McPheeters,R., Meldrin,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Neil,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Pollara,V., Raymond,C., Retta,R.,
            Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J., Rosetti,M.,
            Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Sougnez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Travers,M., Travis,N., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,
            Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAR-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 207418)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kanat,A.,
```

```
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 207418)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,N., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jun 27, 2002 this sequence version replaced gi:21426207.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ----------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ----------- Project Information
            Center project name: L12729
            Center clone name: 344_E_13
FEATURES             Location/Qualifiers
     source          1..207418
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="RP11-344E13"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   1..264
                     /rpt_family="LTR35"
     repeat_region   265..499
                     /rpt_family="L1PA13"
     repeat_region   949..1259
                     /rpt_family="AluY"
     repeat_region   1268..1401
                     /rpt_family="L1MC/D"
     repeat_region   2066..2204
                     /rpt_family="LTR67"
     repeat_region   2210..2340
                     /rpt_family="L1MC/D"
     repeat_region   complement(2412..2546)
                     /rpt_family="L2"
     repeat_region   complement(3163..3459)
                     /rpt_family="L1MC/D"
     repeat_region   6492..6534
                     /rpt_family="AluSq"
     repeat_region   6753..6808
                     /rpt_family="(TG)n"
     repeat_region   complement(7108..7407)
                     /rpt_family="AT_rich"
     repeat_region   7825..8108
                     /rpt_family="AluY"
     repeat_region   complement(8512..8547)
                     /rpt_family="MADE1"
     repeat_region   8875..8940
                     /rpt_family="L2"
     repeat_region   9921..10228
                     /rpt_family="AluSx"
     repeat_region   10713..11010
                     /rpt_family="AluJb"
     repeat_region   12697..12804
                     /rpt_family="FLAM_A"
     repeat_region   complement(12850..13549)
                     /rpt_family="L1PA10"
     repeat_region   13550..13850
                     /rpt_family="AluSp"
     repeat_region   complement(13851..14554)
                     /rpt_family="L1PA10"
     repeat_region   14680..15680
                     /rpt_family="L1PA15"
     repeat_region   15862..15967
                     /rpt_family="L2"
     repeat_region   15968..15988
                     /rpt_family="(TGAA)n"
     repeat_region   16061..17124
                     /rpt_family="L1M4"
     repeat_region   17590..17664
                     /rpt_family="L1M4"
     repeat_region   17665..17967
                     /rpt_family="AluY"
     repeat_region   17968..17979
                     /rpt_family="L1M4"
     repeat_region   17980..18006
                     /rpt_family="(CAAA)n"
     repeat_region   18007..18381
                     /rpt_family="L1M4"
     repeat_region   20417..20580
                     /rpt_family="AluY"
     repeat_region   20586..21111
                     /rpt_family="L1M4"
     repeat_region   complement(21112..21403)
                     /rpt_family="AluSx"
     repeat_region   21404..22015
                     /rpt_family="L1M4"
     repeat_region   22063..22094
                     /rpt_family="AT_rich"
     repeat_region   23322..23380
                     /rpt_family="(TA)n"
     repeat_region   complement(23381..23670)
                     /rpt_family="AluSq"
     repeat_region   complement(25468..26047)
                     /rpt_family="L1ME3A"
     repeat_region   complement(26143..26832)
                     /rpt_family="L1ME3A"
     repeat_region   27610..27727
                     /rpt_family="AluY"
     repeat_region   27728..28199
                     /rpt_family="L1PA4"
     repeat_region   28200..28408
                     /rpt_family="AluY"
     repeat_region   28936..29315
                     /rpt_family="MSTD"
     repeat_region   complement(29697..30168)
                     /rpt_family="L1MC2"
     repeat_region   30281..30372
                     /rpt_family="(TA)n"
```

```
repeat_region    complement(30403..30510)
                 /rpt_family="MIR"
repeat_region    3096..31259
                 /rpt_family="MER44D"
repeat_region    complement(31773..32400)
                 /rpt_family="L2"
repeat_region    32401..32428
                 /rpt_family="AT_rich"

Alignment Scores:
Pred. No.:                                    207418
Score:                7.27e-72    Length:      234
Percent Similarity:   89.89%      Matches:     6
Best Local Similarity: 87.64%     Conservative:
Query Match:          74.42%      Mismatches:  20
DB:                               Indels:      10
                                  Gaps:        3
```

```
Db    59135 ATCTTAAATGTTCATGTCCCA 59155

RESULT 15
AC107926
LOCUS       AC107926    176040 bp    DNA    linear    PRI 01-JUL-2002
DEFINITION  Homo sapiens, clone RP11-381P6, complete sequence.
ACCESSION   AC107926
VERSION     AC107926.6  GI:21637512
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 176040)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-381P6
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 176040)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
            Landers,T., Lehoczky,J., Levine,R., Liu,G., MacLean,C.,
            Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
            McEwan,P., McKernan,K., Meldrim,J., Mereus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
            Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
            Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 176040)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
            Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P. FitzGerald,M., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
            Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
            Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
            Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
            Schupback,R., Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
            Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JUN-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 176040)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
```

```
US-10-054-313-1 (1-286) x AC090774 (1-207418)

Qy      1 MetSerTrpLeuLeuPheLeuValHisArgValAlaLeuAlaAlaLeuProCysArgArg 20
          ||||||||||||||||||||||||  |||||||| |||||| |||||
Db  58364 ATGAGCTGGCTTCTGTTCCTGCACCGC-AGAGTCGGCCGC-AGAGTGGCCGTTGCCCTGC---CGC 58419

Qy     21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyValPhe 40
          ||||  ||||||||||  ||||  ||| |||||||||||||||||||
Db  58420 TGCTCTCACGGGGTTCGGGATGTTCTATGCCCGTGAGGAGGGCGGCAAGACCGGGGTCTTT 58479

Qy     41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
          |||||||||||||||||||||| |||||||||||||||||||||||
Db  58480 CTGACCTGGAATGAGTGCAAAGCACAGGTGGACCCGGTTCCTGCTGCCAGATTTAAGAAG 58539

Qy     61 PheAlaThrGluAspGluValAlaTrpAlaPheValArgGlySerAlaSerProGluValSer 80
          |||||||||||||||  |||||||||||| ||||||||||||||||||||
Db  58540 TTTGCCACAGAGGATGAGAGCCTGGGACTTTGTCAGGAAATCTGCAAGCCCGGAAGTTCA 58599

Qy     81 GluGlyHisGluAsnGlnHisGlyGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
          |||||||||||||||||||||||||||||||||||||||||||
Db  58600 GAAGGGCACGAAGAAAATCAACATGACACAGAATCAGAGACGGCAGCCAGCAAGCGACTCCGT 58659

Qy    101 GluProLeuAspGlyAlaGlyGlyHisGlySerAlaGlnProTyrAlaLysHisMetLysPro 120
          ||||||||  ||||   ||||||||||||||||  ::||||
Db  58660 GAGCCACTGGATGGAGATGGAAGGACGGCAGAGCCATATGC-AAGCACATGAAGCCG 58718

Qy    121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
          |||||  ||  ||||| ||  ||||||||||||||||||||
Db  58719 AGGGTGAAGCCCGCGCCTCCAGTTAGCAGAGACACGTTTCCTACACATGGGAGACTTCGTC 58778

Qy    141 ValValTyrThrAspGlyCysCysSerSerSerAsnGlyArgArgGlyProArgGlyAlaGlyIle 160
          ||||||||||||||||||||||||||||||||  |||||||||||||
Db  58779 GTCGTCTACACGCTGATGGCTGCTGCTCCAGTAATGGGCCGTAGAAGGCCACGAGAGCAGGAAATC 58838

Qy    161 GlyValTyrThrGlyProGlyHisProLeuAsnValGlyLeuIleArgLeuProGlyLysArgGln 180
          ||||||||||||||  ||||||||   :||  |||||||||||||||||||
Db  58839 CGCGGTTTACTGGGGCCCGGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGTGGCAG 58898

Qy    181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaAlaIleGluGlnAlaLysThrGln 200
          |||||||||||||||||||||||||||||||||||||||||||
Db  58899 ACAAACCAAAGAGCGGAAATTCATGCAGCCTACAAGGCCATTGAACAGCAAAGACTCAA 58958

Qy    201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
          |||||||||||||||||||||||||||||||||||||||||
Db  58959 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTGT----AAC 59014

Qy    221 TrpValGlnGlyTrpLysLysGlyAsnGlyTrpLysTyrSerArgAlaGlyLysGluValIleAsn 240
          |||||||||||||||||||||||||||||||||||||||||||
Db  59015 TGGGTTCAAGGTTGGAAGAAAAATGGAAGAACAAGTGCAGGGAAGAGAGGTGATCAAC 59074

Qy    241 LysGluAspPheValAlaLeuGluArgLeuThrGlnLeuGlnMetAspIleGlnTrp------ 258
          |||||||||||||||||||||||||||||||  |||
Db  59075 AAAGAAGACTTCGTGCACTGAGAGGCTGACCCCAGGGGATGGACATTCAGTGGGCGGAGT 59134

Qy    259 --------MetHisValPro 262
          :::|||||||
```

```
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C. Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,R., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (01-JUL-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 1, 2002 this sequence version replaced gi:21427785.
            All repeats were identified using RepeatMasker:
            http://ftp-genome.washington.edu/RM/RepeatMasker.html
            --------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            --------- Project Information
            Center project name: L23134
            Center clone name: 381_P_6
            ---------
            Only the first 176.0 kb of this clone are being submitted.
            The remainder overlaps accession number AC090774 [WICGR project
            L12729].
FEATURES             Location/Qualifiers
     source          1..176040
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-381P6"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   7..233
                     /rpt_family="AluJo"
     repeat_region   477..662
                     /rpt_family="HAL1"
     repeat_region   1499..1785
                     /rpt_family="AluSc"
     repeat_region   1793..1850
                     /rpt_family="L1MB5"
     repeat_region   1851..2131
                     /rpt_family="AluY"
     repeat_region   2135..2163
                     /rpt_family="(CAA)n"
     repeat_region   2164..2936
                     /rpt_family="L1MB5"
     repeat_region   3425..3575
                     /rpt_family="MIR"
     repeat_region   4063..4183
                     /rpt_family="GA-rich"
     repeat_region   4740..4778
                     /rpt_family="AT_rich"
     repeat_region   complement(4785..4940)
                     /rpt_family="FRAM"
     repeat_region   5727..6082
                     /rpt_family="L1MC4a"
     repeat_region   6096..6306
                     /rpt_family="L1MC4a"
     repeat_region   6447..7402
                     /rpt_family="L1MB7"
     repeat_region   7403..7706
                     /rpt_family="AluSq"
     repeat_region   7707..7750
```

```
     repeat_region   7751..8105
                     /rpt_family="L1MB7"
     repeat_region   8106..8680
                     /rpt_family="AluY"
     repeat_region   9419..9524
                     /rpt_family="L1MB7"
     repeat_region   10074..10358
                     /rpt_family="MER103"
     repeat_region   11022..11208
                     /rpt_family="L1MC5"
     repeat_region   11223..11436
                     /rpt_family="L2"
     repeat_region   12191..12212
                     /rpt_family="MIR"
     repeat_region   complement(13126..13321)
                     /rpt_family="(CA)n"
     repeat_region   13920..14143
                     /rpt_family="MLT1E2"
     repeat_region   14198..15784
                     /rpt_family="L2"
     repeat_region   15787..15823
                     /rpt_family="L1PA16"
     repeat_region   complement(16696..16840)
                     /rpt_family="(TG)n"
     repeat_region   complement(17030..17339)
                     /rpt_family="FRAM"
     repeat_region   17575..17802
                     /rpt_family="AluSg"
     repeat_region   17839..17938
                     /rpt_family="L1ME2"
     repeat_region   17939..18247
                     /rpt_family="L1ME2"
     repeat_region   18248..18893
                     /rpt_family="AluSg"
     repeat_region   19076..19398
                     /rpt_family="L1ME2"
     repeat_region   19399..19633
                     /rpt_family="LTR35"
     repeat_region   20083..20093
                     /rpt_family="L1PA13"
     repeat_region   20402..20535
                     /rpt_family="L2"
     repeat_region   21200..21338
                     /rpt_family="L1MC/D"
     repeat_region   21344..21474
                     /rpt_family="LTR67"
     repeat_region   complement(21546..21680)
                     /rpt_family="L1MC/D"
     repeat_region   complement(22297..22593)
                     /rpt_family="L2"
     repeat_region   25626..25660
                     /rpt_family="AluSq"
     repeat_region   25875..25934
                     /rpt_family="(TG)n"
     repeat_region   complement(26234..26533)
                     /rpt_family="AT_rich"
     repeat_region   26951..27234
                     /rpt_family="AluY"
```

```
Alignment Scores:
Pred. No.:                          1.39e-71        Length:               176040
Score:                              1145.50         Matches:              233
Percent Similarity:                 89.51%          Conservative:         6
Best Local Similarity:              87.27%          Mismatches:           21
Query Match:                        74.09%          Indels:               10
DB:                                 9               Gaps:                 3

US-10-054-313-1 (1-286) x AC107926 (1-176040)

Qy      1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20
Db  77485 ATGAGCCCGCTTCTGTTCCTGGCCGC-AGAGTCGCCCTGGCCGCCTTGCCCTGC---CGC 77540
```

```
Qy     21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  77541 TGCTCTCACGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGGCCAAGACCGGGGTCTTT 77600

Qy     41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
          |||||||||||||||||||||| : :||||||||||||||||||||||||||||||||||
Db  77601 CTGACCCGGGAATGAGTGCAAAGCACAGGTGGACCCGGTTTCCTGCTGCCCAGATTAAGGAAG 77660

Qy     61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  77661 TTTGCCACAGAGGATGAGACCTGGGACTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 77720

Qy     81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
          |||||||||||||| |||   ||| |||||||||||||||||||||||||||||||||
Db  77721 GAAGGGCAGGAAAATCAACATGACCAAGAATCAGAGACGAAAGCCAGGCAAGCGACTCCGT 77780

Qy    101 GluProLeuAspGlyHisAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
          ||||||||||||||||| |||||||||  :::|||||||||| -|||||||||||||||
Db  77781 GAGCCACTGGATGGAGATGGAGAGCGCAGAGCCATATGC-AAGCACATGAAAGCCCG 77839

Qy    121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
          ||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  77840 AGGGTGAAGCCGGCCCCTCCAGTAGCGAGAGACACGTTTCCTACATGGGAGACTTCGTC 77899

Qy    141 ValValTyrThrArgProGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
          ||||||||||||||||||||||||||||||| |||:::|||||||||||||||||||||
Db  77900 GTCGTCTACGGTCGATGGCTGCTGCTCCAGTAATGGCGTAGAAGGCCACGAGCAGGAATC 77959

Qy    161 GlyValTyrTrpProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln 180
          ||||||||||||||||||||||: |||||||||||||||||||||||||||||||||
Db  77960 CGGGTTTACTGGGGACCCGGCCTATCCTTTAAATGTAGGCATTAGGACTTCCTGGGTGGCAG 78019

Qy    181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaAlaIleGluGlnAlaLysThrGln 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  78020 ACAAACCAAAGAGCGGAAATTCATGCAGCCTACAGGGCCATTGAACAAGCAAAGACTCAA 78079

Qy    201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
          ||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  78080 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTGT---AAC 78135

Qy    221 TrpValGlnGlyTrpLysAsnGlyTrpTrpSerArgAlaGlyLysGluValIleAsn 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  78136 TGGGTTCAAGGTTGGAAGAAAAATGGATGGAGACAAGTGCAGGGAAAGAGGTGACCAAC 78195

Qy    241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrp------ 258
          |||||||||:::||||||||||||||||||||||||||||||||||||||||||
Db  78196 AAAGAGGACTTCGTGCCACTGGAGAGGCTGACCCAGGGGATGGACATTCAGTGGGCGAGT 78255

Qy    259 -------MetHisValPro 262
                 :::|||||||||
Db  78256 ATCTTAAATGTTCATGTCCCCA 78276
```

OM protein - nucleic search, using frame_plus_p2n model

```
            GenCore version 5.1.6
      Copyright (c) 1993 - 2004 Compugen Ltd.
```

Run on: March 22, 2004, 21:36:56 ; Search time 330 Seconds
(without alignments)
3681.772 Million cell updates/sec

Title: US-10-054-313-1
Perfect score: 1546
Sequence: 1 MSWLLFLAHRVALAALPCR.........FIGNEEADRLAREGAKQSD 286

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop   6.0 , Delext   7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US10054311/runat_22032004_140415_28060/app_query.fasta_1.455
-DB=N_Geneseq_29Jan04 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054313 @CGN_1_352 @runat_22032004_140415_28060 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=10 -THREADS=10 -DSPELEN=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

Result No.	Score	% Match	Query Length	DB	ID	Description
1	1526	98.7	2129	5	ABV30263	Abv30263 Human pro
2	1523	98.5	1150	3	AAZ51265	Aaz51265 Human RNA
3	1522	98.4	965	7	ADA05725	Ada05725 Human NOV
4	1218	78.8	979	9	ADC39107	Adc39107 Novel hum
5	1045.5	67.6	631	7	ADA05719	Ada05719 Human NOV
6	1045.5	67.6	716	7	ADA05717	Ada05717 Human NOV
7	979.5	63.4	586	6	ADA05721	Ada05721 Human NOV
8	905	58.5	545	5	ABV45787	Abv45787 Human pro
9	792	51.2	457	9	ADA05723	Ada05723 Human NOV
10	666.5	43.1	691	5	ABV15988	Abv15988 Human pro
11	606	39.2	519	9	ADD34573	Add34573 Mouse mit
12 c	571	36.9	882	9	ADD34572	Add34572 Mouse mit
13	508.5	32.9	389	3	AAH30241	Aah30241 Human col
14	468	30.3	436	3	ACH42294	Ach42294 Human foe
15	434	28.1	473	8	ABX41815	Abx41815 Bovine ES
16	422	27.3	1071	7	ABL03247	Abl03247 Drosophil
17	399	25.8	407	4	ABX38735	Abx38735 Bovine ES
18	364	23.5	764	6	ABS77198	Abs77198 Frog embr
19	355.5	23.1	3290	4	ABL03246	Abl03246 Drosophil
20 c	272	17.6	119211	4	AAF28553	Aaf28553 Genomic f
21	249.5	16.1	1047	7	ACC61723	Acc61723 Gene sequ
22 c	244	15.8	378	4	AAT24248	Aat24248 Human gen
23	239	15.5	7874	3	AAX20568	Aax20568 Polynucle
24	236.5	15.3	462	5	AAF94380	Aaf94380 Haemophil
25	236.5	15.3	462	4	ABK64944	Abk64944 DNA encod
26	236.5	15.3	1001	4	AAF91449	Aaf91449 Haemophil
27	236.5	15.3	1001	6	ABK37829	Abk37829 DNA seque
28	236.5	15.3	110000	2	AAT42063_01	Continuation (2 of
29	235.5	15.2	2048	2	AAX98244	Aax98244 Contig 13
30	234.5	15.2	435	7	ABZ40958	Abz40958 N. gonorr
31	229.5	14.8	36471	3	AAA81453	Aaa81453 N. mening
32	229.5	14.8	110000	2	AAA81489_2	Continuation (3 of
33	229.5	14.8	349980	3	AAF21611	Aaf21611 Neisseria
34	221.5	14.3	801	3	AAX98246	Aax98246 Nucleotid
35	217	14.0	1509	8	ADA30478	Ada30478 DNA encod
36	213.5	13.8	465	2	AAQ96268	Aaq96268 E.coli ri
37	211.5	13.7	465	2	AAQ96267	Aaq96267 E.coli ri
38	209.5	13.6	757	2	AAQ12780	Aaq12780 Encodes s
39	203	13.1	349980	6	ABQ81849	Abq81849 Bifidobac
40 c	200.5	13.0	471	7	ACF69459	Acf69459 Photorhab
41	200.5	13.0	110000	7	ACF67367_24	Continuation (25 o
42 c	200.5	13.0	110000	5	ACF65386_4	Continuation (5 of
43	190.5	12.3	630	8	ADB11347	Adb11347 Alloiococ
44	190.5	12.3	630	8	ADB11349	Adb11349 Alloiococ
45	190.5	12.3	110000	8	ADB12064_15	Continuation (16 o

ALIGNMENTS

RESULT 1
ABV30263
ID   ABV30263 standard; cDNA; 2129 BP.
AC   ABV30263;
XX
DT   16-SEP-2002 (first entry)
XX
DE   Human prostate expression marker cDNA 30254.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US005171.
XX
PR   17-FEB-2000; 2000US-0183319P.
PR   16-MAR-2000; 2000US-0189862P.
PR   25-MAY-2000; 2000US-0207454P.
PR   09-JUN-2000; 2000US-0211314P.
PR   18-JUL-2000; 2000US-0219007P.
PR   13-DEC-2000; 2000US-0255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,   Endege WO, Monahan JE;
XX

```
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS   Claim 1; Page 6567-6568; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for: (a) assessing whether
CC   a patient is afflicted with prostate cancer; (b) monitoring the
CC   progression of prostate cancer in a patient; (c) assessing the efficacy
CC   of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC   the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC   determining whether prostate cancer has metastasized in a patient; (h)
CC   assessing the aggressiveness or indolence of prostate cancer in a patient
CC   ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ   Sequence 2129 BP; 609 A; 510 C; 578 G; 427 T; 0 U; 5 Other;

Alignment Scores:
Pred. No.:
Score:                   1526.00          Length:          2129
Percent Similarity:      99.30%           Matches:          282
Best Local Similarity:   98.60%           Conservative:       2
Query Match:             98.71%           Mismatches:         2
DB:                      5                Indels:             0
                         1.5e-125         Gaps:               0

US-10-054-313-1 (1-286) x ABV30263 (1-2129)

Qy    1 MetSerTrpLeuPheLeuAlaHisArgValAlaAlaAlaLeuAlaAlaLeuProCysArgArg  20
Db  112 ATGAGCTGGCTTCGTTCCTGGCCCACAGAGTCGCCTTGCCCGCCTTGCCCTGCCCGCCGC   171

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgLysThrGlyValPhe           40
Db  172 GGCTCTCGCGGGGTTCGGGGATGTTCTATGCCGTGAGGAGGGGCCCGAAGACCGGGGTCTTT  231

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys    60
Db  232 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTCCTGCTGCCAGATTTAAGAAG     291

Qy   61 PheAlaThrGluAspGluAlaTrpAlaProPheValArgLysSerAlaSerProGluValSer  80
Db  292 TTTGCCACAGAGGATGAGGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA   351

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg    100
Db  352 GAAGGGCATGAAAATCAACATGGACCAAGAATCGGAGCGGAAGCCCAGCAAGGACTCCGT    411

Qy  101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro    120
Db  412 GAGCCACTGGATGGAGATGACATGAAAGCCGCAGAGCCGTATGCAAAGCACATGAAGCCG    471

Qy  121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
Db  472 AGCGTGGAGCCGGCCGCCCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC    531

Qy  141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgProArgAlaGlyIle 160
Db  532 GTCGTTCACACTGATGGCTGCTGCTCCAGTAATGGCCGTAGAAGGCCGGAGCAGGAATC    591

Qy  161 GlyValThrTrpGlyTyrProGlyHisProLeuAsnValGlyIleLeuProGlyArgGln    180
Db  592 GGCGTTTACTGGGGCCAGCCATCTTAAATGTAGGCATTAGGACTTCCTGGGCGCAG       651

Qy  181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysLysAlaIleGluGlnAlaLysThrGln 200
Db  652 ACAAACCAAAGAGCGGAAATCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA    711


Qy  201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn   220
Db  712 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC   771

Qy  221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn   240
Db  772 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC   831

Qy  241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleTrpGlnTrpMetHis 260
Db  832 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT   891

Qy  261 ValProGlyHisSerGlyPheIleGlyAsnGluAlaArgLeuAlaArgGluGly         280
Db  892 GTTCCTGGTCATTCGGGGATTTATAGGCAATGAAGCTGACAGATTAGCCAGAGAGGA     951

Qy  281 AlaLysGlnSerGluAsp 286
Db  952 GCTAAACAATCCGAAGAC 969


RESULT 2
AAZ51265
ID   AAZ51265 standard; cDNA; 1150 BP.
XX
AC   AAZ51265;
XX
DT   06-JUN-2000 (first entry)
XX
DE   Human RNA-associated protein-16 (RNAAP-16) encoding cDNA.
XX
KW   RNA-associated protein; RNAAP; human; clone 2073417; cytostatic;
KW   immunosuppressive; antiinflammatory; antiarthritic; keratolytic; neuroprotective;
KW   antiarteriosclerotic; hepatotropic; antipsoriatic; virucide; anti-HIV;
KW   antiallergic; antirheumatic; anticarthritic; opthalmological; autoimmune;
KW   antimicrobial; cell proliferative disorder; inflammation; cirrhosis;
KW   actinic keratosis; bursitis; arteriosclerosis; artherosclerosis;
KW   hepatitis; myelofibrosis; primary thrombocythemia; psoriasis; cancer;
KW   mixed connective tissue disease; MCTD; HIV; uveitis; Crohn's disease;
KW   allergy; rheumatoid arthritis; parasitic infection; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             85..945
FT                   /*tag= a
FT                   /product= "Human RNA-associated protein-16"
FT                   /note= "Derived from ISLTNOT01 library"
FT   sig_peptide     85..138
FT                   /*tag= b
FT   mat_peptide     139..942
FT                   /*tag= c
FT                   /product= "Mature human RNA-associated protein-16"
XX
PN   WO20011171-A2.
XX
PD   02-MAR-2000.
XX
PF   20-AUG-1999;    99WO-US019361.
XX
PR   21-AUG-1998;    98US-0097550P.
PR   12-JAN-1999;    99US-0115639P.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Hillman JL,  Yue H,  Tang YT,  Corley NC,  Guegler KJ,  Gorgone GA;
PI   Patterson C,  Baughn MR,  Lal P,  Bandman O,  Reddy R,  Azimzai Y;
PI   Shih LL,  Yang J,  Lu DAM;
XX
DR   WPI; 2000-237651/20.
DR   P-PSDB; AAY70235.
XX
PT   Human RNA-associated proteins useful in diagnosing, treating and
PT   preventing cell proliferative, autoimmune, inflammatory and infectious
```

```
PT    disorders.
XX
PS    Claim 9; Page 117-118; 123pp; English.
XX
CC    The present sequence is the cDNA encoding human RNA-associated protein-16
CC    (RNAAP-16), identified in Incyte clone 2073417, derived from ISLTNOT01
CC    library. It is expressed in nervous, reproductive, gastrointestinal,
CC    cardiovascular and haematopoietic/immune tissues. It has cytostatic,
CC    immunosuppressive, antiinflammatory, antiarteriosclerotic, hepatotropic,
CC    keratolytic, neuroprotective, antipsoriatic, anti-HIV, antiallergic,
CC    antirheumatic, virucide, antiarthritic, opthalmological and antimicrobial
CC    activity. RNAAP antibodies are useful for diagnosis of diseases
CC    associated with altered expression or activity of RNAAP. It is used to
CC    treat cell proliferative, autoimmune, inflammatory and infectious
CC    disorders, like actinic keratosis, bursitis, arteriosclerosis,
CC    artheroscierosis, cirrhosis, hepatitis, myelofibrosis, mixed connective
CC    tissue disease (MCTD), psoriasis, primary thrombocythemia and cancer,
CC    HIV, allergies, rheumatoid arthritis, uveitis, Crohn's disease, and
CC    bacterial, viral and parasitic infections
XX
SQ    Sequence 1150 BP; 302 A; 250 C; 352 G; 246 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:              1.3e-125     Length:          1150
Score:                  1523.00      Matches:          281
Percent Similarity:     99.30%       Conservative:       3
Best Local Similarity:  98.25%       Mismatches:         2
Query Match:            98.51%       Indels:             0
DB:                     3            Gaps:               0

US-10-054-313-1 (1-286) x AA251265 (1-1150)

Qy      1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaAlaAlaAlaLeuProCysArgArg  20
Db     85 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCGCCGCTGCCCTGCCGCGGCGC  144

Qy     21 GlySerArgArgGlyPheGlyMetPheTyrAlaValArgArgLysThrGlyValPhe  40
Db    145 GGCTCTCGCGGGGTTCGGGATGTTCTATGCCGTGAGGAGGAAGACCGGGGTCTTT  204

Qy     41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys  60
Db    205 CTGACCTGGAATGAGTGCAGAGCACAGGTTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG  264

Qy     61 PheAlaThrGluAspGluAlaProProValSerProAlaPheValArgLysSerAlaSerProGluValSer  80
Db    265 TTTGCCACAGAGGATGAGGCCTGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA  324

Qy     81 GluGlyHisGluAsnGlnHisGlyValGlnGluSerGluAlaGlyLysArgGlyLeuArg  100
Db    325 GAAGGGCATGAAAAATCAACATGGACAAGAATCGGAAGCGGAAGCGACTCCCT  384

Qy    101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
Db    385 GAGCCACTGGATGGAGATGGACATGAAAGCCAGAGCCGTATGCAAAGCACATGAAGCCG  444

Qy    121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal  140
Db    445 AGCATGGAGCCGGCGCCGCCTGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC  504

Qy    141 ValValTyrThrArgGlyTyrCysCysSerSerAsnGlyArgArgLysProAlaGlyIle  160
Db    505 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCCGCAGCAGGAATC  564

Qy    161 GlyValTyrTrpGlyTyrProGlyHisProLeuAsnValGlyIleLeuArgLeuProGlyArgGln  180
Db    565 GGCGTTTACTGGGGCGGGCCAGGCCCATCCTTAAATGTAGGCATTAGACTTCCTGGGCGCCAG  624

Qy    181 ThrAsnGlnArgAlaGlyIleHisAlaAlaCysLysAlaIleGluGlnLnAlaLysThrGln  200
Db    625 ACAAACCAAAGAGCCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAGACTCAA  684

Qy    201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleIleThrAsn  220
```

```
Db    685 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC  744

Qy    221 TrpValGlnGlyTyrLysAsnGlyTrpLysSerAlaGlyTyrThrSerAlaGlyLysValGluValIleAsn  240
Db    745 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGAGCAGTGCAGGGAAGAGAGGTGATCAAC  804

Qy    241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyValMetAspIleGlnTrpProMetHis  260
Db    805 AAAGAGGACTTTGTGGCACTGGAGAGCGCTTACCAGGGATGGACATTCAGTGGAGTGCAT  864

Qy    261 ValProGlyHisSerGlyPheIleGlyPheIleGlyAsnGluAlaAspArgLeuAlaArgGluGlyGly  280
Db    865 GTTCCTGGTCATTCGGGATTTATAGGCAATGGCAATGAGGCTGACAGATTAGCCAGAGAAGGA  924

Qy    281 AlaLysGlnSerGluAsp  286
Db    925 GCTAAACAATCCGAAGAC  942
```

```
RESULT 3
ADA05725
ID   ADA05725  standard; cDNA; 965 BP.
XX
AC   ADA05725;
XX
DT   06-NOV-2003  (first entry)
XX
DE   Human NOV15e encoding cDNA SEQ ID NO:85.
XX
KW   human; NOVX; antidiabetic; anorectic; antibacterial; virucide;
KW   immunomodulator; cytostatic; nootropic; neuroprotective;
KW   antiparkinsonian; antilipaemic; gene therapy; human disease;
KW   metabolic disorder; diabetes; obesity; infection; cachexia; cancer;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia; gene; ss.
XX
OS   Homo sapiens.
XX
FH                  Location/Qualifiers
FT   Key
FT   CDS             10..870
FT                   /*tag= a
FT                   /product= "NOV15e"
XX
PN   WO2003029424-A2.
XX
PD   10-APR-2003.
XX
PF   02-OCT-2002; 2002WO-US031373.
XX
PR   02-OCT-2001; 2001US-0326483P.
PR   05-OCT-2001; 2001US-0327435P.
PR   05-OCT-2001; 2001US-0327499P.
PR   09-OCT-2001; 2001US-0327917P.
PR   09-OCT-2001; 2001US-0328029P.
PR   09-OCT-2001; 2001US-0328044P.
PR   09-OCT-2001; 2001US-0328056P.
PR   12-OCT-2001; 2001US-0328849P.
PR   15-OCT-2001; 2001US-0329414P.
PR   17-OCT-2001; 2001US-0330142P.
PR   18-OCT-2001; 2001US-0303099P.
PR   22-OCT-2001; 2001US-0341058P.
PR   24-OCT-2001; 2001US-0339266P.
PR   24-OCT-2001; 2001US-0343629P.
PR   29-OCT-2001; 2001US-0349575P.
PR   01-NOV-2001; 2001US-0363577P.
PR   17-APR-2002; 2002US-0373260P.
PR   19-APR-2002; 2002US-0373815P.
PR   19-APR-2002; 2002US-0373817P.
PR   19-APR-2002; 2002US-0373826P.
PR   19-APR-2002; 2002US-0373884P.
PR   22-APR-2002; 2002US-0374977P.
PR   16-MAY-2002; 2002US-0381037P.
PR   16-MAY-2002; 2002US-0381038P.
```

```
PR   16-MAY-2002; 2002US-0381042P.
PR   17-MAY-2002; 2002US-0381642P.
PR   28-MAY-2002; 2002US-0383656P.
PR   29-MAY-2002; 2002US-0383831P.
PR   25-JUN-2002; 2002US-0391335P.
PR   01-OCT-2002; 2002US-00262511.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;
PI   Paturajan M, Spytek KA, Edinger SR, Ellerman K, Malvankar UM;
PI   Ort T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Catterton E;
PI   Ji W, Miller CE, Rastelli L, Stone DJ, Pena CEA, Shenoy SG;
PI   Shinkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;
PI   Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;
XX
DR   WPI; 2003-381626/36.
DR   P-PSDB; ADA05726.
XX
PT   New NOVX polypeptides and nucleic acids, useful for diagnosing,
PT   preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,
PT   cancer or dyslipidemia, and in chromosome mapping, tissue typing or
PT   pharmacogenomics.
XX
PS   Claim 20; Page 160; Page 586pp; English.
XX
CC   The present invention describes NOVX proteins, where X can be 1 to 55
CC   (e.g. NOV1). Also described: (1) a composition comprising a polypeptide
CC   described above and a carrier; (2) a kit comprising, in one or more
CC   containers, the composition described above; (3) an isolated nucleic acid
CC   molecule which encodes a NOVX protein of the invention; (4) a vector
CC   comprising the nucleic acid molecule described above; (5) a cell
CC   comprising the above vector; (6) an antibody that immunospecifically
CC   binds to the polypeptide described above; (7) methods for determining the
CC   presence or amount of the above polypeptide or nucleic acid molecule in a
CC   sample; (8) methods for determining the presence of or predisposition to
CC   a disease associated with altered levels of expression of the above
CC   polypeptide or nucleic acid molecule in a first mammalian subject; (9) a
CC   method of identifying an agent that binds to the polypeptide described
CC   above; (10) a method for identifying a potential therapeutic agent for
CC   use in treating a pathology that is related to an aberrant expression or
CC   aberrant physiological interactions of the polypeptide; (11) a method of
CC   screening for a modulator of activity or of latency or predisposition to
CC   a pathology associated with the polypeptide; (12) a method for modulating
CC   the activity of the polypeptide described above; (13) methods of treating
CC   or preventing a pathology associated with the above polypeptide in a
CC   mammal; and (14) a method for producing the above polypeptide. NOVX
CC   sequences have antidiabetic, anorectic, antibacterial, virucide,
CC   immunomodulator, cytostatic, nootropic, neuroprotective, antiparkinsonian
CC   and antilipaemic activities, and can be used in gene therapy. The
CC   polypeptide is useful in manufacturing a medicament for treating a
CC   syndrome associated with a human disease. The polypeptide or the nucleic
CC   acid molecule may be used to diagnose, treat or prevent metabolic
CC   disorders such as diabetes or obesity, infections, cachexia, cancer,
CC   neurodegenerative disorders such as Alzheimer's disease or Parkinson's
CC   disease, immune disorders, haematopoietic disorders and various
CC   dyslipidaemias. The nucleic acids can also be used as hybridisation
CC   probes, in chromosome mapping, tissue typing, preventive medicine and
CC   pharmacogenomics. The present sequence encodes a human NOVX protein from
CC   the present invention.
XX
SQ   Sequence 965 BP; 258 A; 210 C; 295 G; 202 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score:                      1522.00
Percent Similarity:          98.95%
Best Local Similarity:       98.25%
Query Match:                 98.45%
DB:

Length:          965
Matches:         281
Conservative:      2
Mismatches:        3
Indels:            0
Gaps:              0

                 1.28e-125

US-10-054-313-1 (1-286) x ADA05725 (1-965)
```

```
Qy   1   MetSerTrpLeuLeuPheLeuAlaHisArgValAlaAlaAlaAlaLeuProCysArgArg   20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10   ATGAGCTGGTTGTTCTTGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCTTGCCCTGCCGC   69

Qy  21   GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe   40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  70   GGCTCTCCGGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT  129

Qy  41   LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys   60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 130   CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG  189

Qy  61   PheAlaThrGluAspGluAlaTrpAlaSerProGluValSer                     80
             ||||||||||||||||||||||||||||||||||||||||||
Db 190   TTTGCCACAGAGGATGAGGCCTGGGCCTTGTCAGGAAATCTGCAAGCCCGGAAGTTCA   249

Qy  81   GluGlyHisGluAsnGlnHisGlyGlnLeuSerGluAlaLysProGlyLysArgLeuArg  100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 250   GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGCGAAAGCCAGCAAGCGACTCCGT  309

Qy 101   GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
             ||||||||||||||||||||||||||||||||||| :: :||||||||||||||||||||
Db 310   GAGCCACTGGATGGAGATGGACATGAAAGCGCAGACGCGTATGCAAAGCACATGAAGCCG  369

Qy 121   SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal  140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 370   AGCGTGGAGCCCGCGCCCTCCAGTTAGGACGAGAGACACGTTTTCCTACATGGAGACTTCGTC  429

Qy 141   ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgGlyProArgAlaGlyIle  160
             |||||||||||||||||||||||||||||::: ||||||||||||||||||||||||||||
Db 430   GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGGAGCCAGGAATC  489

Qy 161   GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgProLeuProArgGln  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 490   GGCGTTTACTGGGGCCCAGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGCCAG  549

Qy 181   ThrAsnGlnArgAlaGlyIleHisAlaAlaCyBLysAlaAlleGluGlnAlaLysThrGln  200
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 550   ACAAACCAGAGGGCAGAAATTCATGCAGCCTGCAAAGCCATTGACACAGCAAAGACTCAA  609

Qy 201   AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 610   AACATCAATAAACTGGTTCTGTATCACAGACAGTATGTTTACGATAAATGGTATAACTAAC  669

Qy 221   TrpValGlnGlyTrpLysLysAsnGlyTrpThrGlyGlyLysGluValIleAsn       240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 670   TGGGTTCAGGTTGGAAGAAAAATTGGGTGGAAGACAAGTGCAGGGAAGAGGTGATCAAC  729

Qy 241   LysGluAspPheValAlaLeuGluThrGlnGluMetAspIleGlnTrpMetHis       260
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 730   AAAGAGGACTTTGTGGCACTGGAGACGGCCTTACCCAGGGATGGACATTCAGTGGATGCAT  789

Qy 261   ValProGlyHisSerGlyPheIleGlyAsnGluGluUlaAlaArgLeuAlaArgGluGly  280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 790   GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGCACGAGATTAGCCAGAGAAGGA  849

Qy 281   AlaLysGlnSerGluAsp                                            286
             ||||||||||||||||||
Db 850   GCTAAACAATCGGAGGAC                                            867
```

```
RESULT 4
ADC39107
ID   ADC39107 standard; cDNA; 979 BP.
XX
AC   ADC39107;
XX
DT   18-DEC-2003 (first entry)
XX
DE   Novel human NOVX polypeptide coding sequence SEQ ID NO: 49.
XX
KW   ds; gene; antidiabetic; cytostatic; immunomodulator; anorectic;
KW   antilipemic; nootropic; neuroprotective; immunostimulant;
KW   antiparkinsonian; anti-HIV; antiasthmatic; antiinflammatory; hypotensive;
```

KW antiarteriosclerotic; hemostatic; osteopathic; gene therapy.; NOVX;
KW diabetes; obesity; cancer; lymphoma; uterus cancer; prostate cancer;
KW dyslipidemia; anorexia; wasting disorder; Alzheimer's disease;
KW Parkinson's disorder; cachexia; cardiomyopathy; AIDS; asthma;
KW Crohn's disease; multiple sclerosis; hypertension; atherosclerosis;
KW hemophilia; graft-versus-host disease;
KW Albright hereditary osteodystrophy.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT CDS   90..897
FT   /*tag= a
XX
PN WO200301O327-A2.
XX
PD 06-FEB-2003.
XX
PF 02-MAY-2002; 2002WO-US014199.
XX
PR 02-MAY-2002; 2002WO-US014199.
PR 03-MAY-2001; 2001US-0288063P.
PR 03-MAY-2001; 2001US-0288395P.
PR 07-MAY-2001; 2001US-0289087P.
PR 09-MAY-2001; 2001US-0289817P.
PR 09-MAY-2001; 2001US-0289818P.
PR 11-MAY-2001; 2001US-0290194P.
PR 14-MAY-2001; 2001US-0290753P.
PR 15-MAY-2001; 2001US-0291181P.
PR 16-MAY-2001; 2001US-0291243P.
PR 18-MAY-2001; 2001US-0292001P.
PR 21-MAY-2001; 2001US-0292374P.
PR 22-MAY-2001; 2001US-0292587P.
PR 23-MAY-2001; 2001US-0293107P.
PR 25-MAY-2001; 2001US-0293747P.
PR 29-MAY-2001; 2001US-0294109P.
PR 29-MAY-2001; 2001US-0294110P.
PR 30-MAY-2001; 2001US-0294434P.
PR 31-MAY-2001; 2001US-0294827P.
PR 31-JUL-2001; 2001US-0304879P.
PR 31-JUL-2001; 2001US-0309194P.
PR 14-AUG-2001; 2001US-0312270P.
PR 17-AUG-2001; 2001US-0313416P.
PR 10-SEP-2001; 2001US-0318463P.
PR 27-SEP-2001; 2001US-0325683P.
PR 18-OCT-2001; 2001US-0330292P.
PR 28-NOV-2001; 2001US-0333873P.
PR 03-DEC-2001; 2001US-0336909P.
PR 03-DEC-2001; 2001US-0337552P.
PR 21-FEB-2002; 2002US-0059245P.
PR 01-MAY-2002; 2002US-0136826.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Miller CE, Kekuda R, Malvankar UM, Li L, Pena CEA, Spytek KA;
PI Gorman L, Guo X, Fernandes ER, Smithson G, Stone DJ, Zerhusen BD;
PI Patturajan M, Anderson DW, Mezes PS, Peyman JA, Macdougall JR;
PI Padigaru M, Rastelli L, Shenoy SG, Gerlach VL, Shimkets RA, Zhong M;
PI Edinger SR, Ellerman K;
XX
DR WPI; 2003-239445/23.
DR P-PSDB; ADC39108.
XX
PT New NOVX polypeptides and polynucleotides, useful in gene therapy,
PT particularly for treating or preventing a syndrome associated with a
PT human disease e.g. diabetes, obesity, cancer, Alzheimer's disease,
PT hypertension or hemophilia.
XX
PS Claim 20; SEQ ID NO 49; 748pp; English.
XX
CC The invention relates to new isolated NOVX polypeptides, the genes
CC encoding them or sequences having at least 95% identity to the amino acid
CC or nucleotide sequences. The NOVX polypeptide is useful as a therapeutic,
CC particularly in the manufacture of a medicament for treating a syndrome

CC associated with a human disease, which includes a pathology associated
CC with NOVX polypeptide. The NOVX polypeptide is particularly useful for
CC treating, preventing or alleviating pathology associated with NOVX
CC polypeptide in a mammal, e.g. a human. The NOVX nucleic acid and
CC polypeptide are especially useful for treating or preventing e.g.
CC diabetes, obesity, cancers (e.g. lymphoma, uterus cancer or prostate
CC cancer), dyslipidemias, anorexia, cachexia, cardiomyopathy, AIDS, asthma, Crohn's
CC Parkinson's disorder, cachexia, cardiomyopathy, AIDS, asthma, Crohn's
CC disease, multiple sclerosis, hypertension, atherosclerosis, hemophilia,
CC graft-versus-host disease or Albright hereditary osteodystrophy. The DNA
CC encoding the protein is useful in gene therapy for treating the above
CC conditions. These are also useful in developing powerful assay system for
CC functional analysis of various human disorders, as well as in diagnostic
CC applications. This sequence represents one of the NOVX genes of the
CC invention.
XX
SQ Sequence 979 BP; 268 A; 205 C; 286 G; 220 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   1.16e-98
Score: 1218.00
Percent Similarity: 90.64%
Best Local Similarity: 88.01%
Query Match: 78.78%
DB:

Length: 979
Matches: 235
Conservative: 7
Mismatches: 19
Indels: 6
Gaps: 2

US-10-054-313-1 (1-286) x ADC39107 (1-979)

Qy   1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   90 ATGAGCTGGCTGTTCCTGCCCGGCGGCAGAGTCGGCTTGGCCGCCCTTGCCCTGCGCCGC 149

Qy   21 GlySerArgGlyTyrPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe 40
         |||||||||||||||||||||||||||||||||||||||||||| ||:::: |||||||||
Db 150 ---TCTCGCGGGTTCGGGATGTTCATGCCGTGAGGAGGGCGGGAAGTCCGCGGTCTTT 206

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
         |||||||||||:::||||||||||||||||||||||||||||||||||||||||||||
Db 207 CTGACCCGGAATGAGTGCAAGCACAGGTGAACCGGTTCCTGCTGCTGCCAGATTAAGAAG 266

Qy   61 PheAlaThrGluAspGluAlaTrpAlaTrpAlaPheValArgLysSerProGluValSer 80
         |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db 267 TTTGCCACAGAGGATGAGGCCTGGACTTTGTCAGGAAATCTGCAGAGCCCGGAAGTTTCA 326

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGlnAlaLysProGlyLysArgLeuArg 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 327 GAAGGGCAGGAAAATCAACATGGACAAGAATCGGAGACGAAAGCCAAGCGACTCCGT 386

Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
         |||||||||||||||||||||||||||||||||||||:::||||||||||||||||
Db 387 GAGCCACTGGATGGAGATGGAGATGGAGCAGAGCCATATGCAAGCACAATGAAGCCG 446

Qy 121 SerValGluProAlaProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
         |||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db 447 AGCGTGAAGCCGGCCGCCTCCAGTTAGCAGAGACACGTTTCTCACATGGGAGACTTCGTC 506

Qy 141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgGlyLysProArgAlaGlyIle 160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 507 GTCGTCTACGCTCACGGCTGCTGCTCCAGCAATGGGCGTAGGAAGCCACGAGCAGGAAATC 566

Qy 161 GlyValTyrTrpProGlyTyrProGlyHisProLeuAsnValGlyIleArgGlyLeuProGlyArgGln 180
         |||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db 567 CGCGTTACTGGGGGCCCGGGCTATCCTTAACTGTAGGCATTAGACTTCCTGGGTGCAG 626

Qy 181 ThrArgAsnGlnArgAlaGluIleHisAlaAlaCysAlaIleGluGlnAlaLysThrGln 200
         |||||||||||||||||||||||||||||||||||||||||||||:::|||||||||
Db 627 ACAAACCAAGAGCGAAATTCATGCAGCCTGCAAGGCCATTGACCAAGCAGACTCAG 686

Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
         |||||||||||||||||||||||||::::||||||||||||||||||||||||||
Db 687 AAAATCAATAAACTGGTTCGTATACGACAGTATGTTTACGATAAATGGTATAACTAAC 746

Qy 221 TrpValGlnIleGlyTrpLysAsnGlyTrpLysThrSerAlaGlyValIleGluValIleAsn 240

```
Db    747  TGGGTTCAAGGTTGGAAGAAAATGGGTGGAAGACAAGTGCAGGGAAGAGGTGATCAAC  806
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    241  LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrp------  258
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db    807  AAAGAGGACTTCGTGGCACTGGAGAGGCTGACCCAGGGGATGGACATTCAGTGGGCGAGT  866

Qy    259  --------MetHisValPro  262
                   ::|||||||||
Db    867  ATCTTAAATGTTCATGTCCCA  887
```

RESULT 5
ADA05719
ID   ADA05719 standard; cDNA; 631 BP.
XX
AC   ADA05719;
XX
DT   06-NOV-2003 (first entry)
XX
DE   Human NOV15b encoding cDNA SEQ ID NO:79.
XX
KW   human; NOVX; antidiabetic; anorectic; antibacterial; virucide;
KW   immunomodulator; cytostatic; nootropic; neuroprotective;
KW   antiparkinsonian; antilipaemic; gene therapy; human disease;
KW   metabolic disorder; diabetes; obesity; infection; cachexia; cancer;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   2..631
FT   /tag=   a
FT   /product= "NOV15b"
XX
PN   WO2003029424-A2.
XX
PD   10-APR-2003.
XX
PF   02-OCT-2002; 2002WO-US031373.
XX
PR   02-OCT-2002; 2001US-0349575P.
PR   29-OCT-2001; 2001US-0346357P.
PR   01-NOV-2001; 2001US-0373260P.
PR   17-APR-2002; 2002US-0373815P.
PR   19-APR-2002; 2002US-0327449P.
PR   09-OCT-2001; 2001US-0327917P.
PR   09-OCT-2001; 2001US-0373817P.
PR   19-APR-2002; 2002US-0373826P.
PR   19-APR-2002; 2002US-0328029P.
PR   09-OCT-2001; 2001US-0328044P.
PR   22-APR-2002; 2002US-0373884P.
PR   09-OCT-2001; 2001US-0328056P.
PR   16-MAY-2002; 2002US-0374977P.
PR   12-OCT-2001; 2001US-0328849P.
PR   16-MAY-2002; 2002US-0381037P.
PR   15-OCT-2001; 2001US-0329414P.
PR   16-MAY-2002; 2002US-0381038P.
PR   17-OCT-2001; 2001US-0330142P.
PR   17-MAY-2002; 2002US-0381042P.
PR   18-OCT-2001; 2001US-0330309P.
PR   28-MAY-2002; 2002US-0383556P.
PR   22-OCT-2001; 2001US-0341058P.
PR   29-MAY-2002; 2002US-0383831P.
PR   24-OCT-2001; 2001US-0373884P.
PR   25-JUN-2002; 2002US-0391335P.
PR   24-OCT-2001; 2001US-0343629P.
PR   01-OCT-2002; 2002US-0262511.
XX
PA   (CURA-) CURAGEN CORP.
```

```
Db         Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;
PI         Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;
PI         Ort T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Catterton E;
PI         Ji W, Miller CE, Rastelli L, Stone DJ, Pena CEA, Shenoy SG;
PI         Shimkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;
PI         Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;
XX
DR         WPI; 2003-381626/36.
DR         P-PSDB; ADA05720.
XX
PT         New NOVX polypeptides and nucleic acids, useful for diagnosing,
PT         preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,
PT         cancer or dyslipidaemia, and in chromosome mapping, tissue typing or
PT         pharmacogenomics.
XX
PS         Claim 20; Page 159; 586pp; English.
XX
CC         The present invention describes NOVX proteins, where X can be 1 to 55
CC         (e.g. NOV1). Also described: (1) a composition comprising a polypeptide
CC         described above and a carrier; (2) a kit comprising, in one or more
CC         containers, the composition described above; (3) an isolated nucleic acid
CC         molecule which encodes a NOVX protein of the invention; (4) a vector
CC         comprising the nucleic acid molecule described above; (5) a cell
CC         comprising the above vector; (6) an antibody that immunospecifically
CC         binds to the polypeptide described above; (7) methods for determining the
CC         presence or amount of the above polypeptide or nucleic acid molecule in a
CC         sample; (8) methods for determining the presence of or predisposition to
CC         a disease associated with altered levels of expression of the above
CC         polypeptide or nucleic acid molecule in a first mammalian subject; (9) a
CC         method of identifying an agent that binds to the polypeptide described
CC         above; (10) a method for identifying a potential therapeutic agent for
CC         use in treating a pathology that is related to an aberrant expression or
CC         aberrant physiological interactions of the polypeptide; (11) a method of
CC         screening for a modulator of activity or of latency or predisposition to
CC         a pathology associated with the polypeptide; (12) a method for modulating
CC         the activity of the polypeptide described above; (13) methods of treating
CC         or preventing a pathology associated with the above polypeptide. NOVX
CC         mammal; and (14) a method for producing the above polypeptide. NOVX
CC         sequences have antidiabetic, anorectic, antibacterial, virucide,
CC         immunomodulator, cytostatic, nootropic, neuroprotective, antiparkinsonian
CC         and antilipaemic activities, and can be used in gene therapy. The
CC         polypeptide is useful in manufacturing a medicament for treating a
CC         syndrome associated with a human disease. The polypeptide or the nucleic
CC         acid molecule may be used to diagnose, treat or prevent metabolic
CC         disorders such as diabetes or obesity, infections, cachexia, cancer,
CC         neurodegenerative disorders such as Alzheimer's disease or Parkinson's
CC         disease, immune disorders, haematopoietic disorders and various
CC         dyslipidaemias. The nucleic acids can also be used as hybridisation
CC         probes, in chromosome mapping, tissue typing, preventive medicine and
CC         pharmacogenomics. The present sequence encodes a human NOVX protein from
CC         the present invention.
XX
SQ         Sequence 631 BP; 167 A; 140 C; 189 G; 135 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                            631
Score:                     1.32e-83   201
Percent Similarity:        1045.50    631
Best Local Similarity:     70.63%     Conservative:  1
Query Match:               70.28%     Mismatches:    1
DB:                        67.63%     Indels:        83
                           7          Gaps:          1
```

```
US-10-054-313-1 (1-286)  x ADA05719 (1-631)

Qy     1  MetSerTrpLeuLeuPheLeuAlaHisArgValAlaAlaAlaLeuProCysArgArg  20
                                                               ||||||
Db    14  ATGAGCTGGTTGTTCTGTTCCTGGCCCACAGAGTCGGCTTGCCGCCTTGCCGCCGC  73

Qy    21  GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgGlyLysThrGlyValPhe  40
                                                                  ||||||
Db    74  GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCCAAGACCGGGGTCTTT  133
```

```
Qy   41  LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys  60
             |||||||||||||||||||||||||||||||||||
Db   134 CTGACCTGGAATGAGTGC-----------------------------------------  151

Qy   61  PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer  80
             |||||||||||||
Db   151 -----------------------------------------------------------  151

Qy   81  GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg  100
             |||||||||||||||
Db   151 -----------------------------------------------------------  151

Qy   101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
             |||||||||||||||||
Db   151 -----------------------------------------------------------  151

Qy   121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
                                            ||||||||||||||||||||||
Db   152 --------------AGAGACACGTTTTCCTACATGGGAGACTTCGTC              184

Qy   141 ValValTyrThrAspGlyCysCysSerSerAsnGlyTyrArgGlyLysProAlaGlyIle  160
             ||||||||||||||||||||||||||||| ||| |||
Db   185 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGGTAGAGGGCCGCGAGCAGGAAATC  244

Qy   161 GlyValTyrTrpProGlyHisProLeuAsnValGlyIleArgGlyProGlyLyArgGln  180
             |||||||||||||||||||||||||||||||||||||||
Db   245 GGGGTTTACTGGGGGGCCGGGCCATCCTTAAATGTAGGCATTAGACTTCCTGGGCCGCAG  304

Qy   181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln  200
             |||||||||||||||||||||||||||||||||||||||||||
Db   305 ACAAACCAAAGAGCGGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA  364

Qy   201 AsnIleAsnLysLeuValLeuThrTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
             |||||||||||||||||||||||||||||||||||||
Db   365 AACATCAATAAACTGGTTCTGATACAGACAGTATGTTTACCATAAATGGTATAACTAAC  424

Qy   221 TrpValGlnGlyTyrTrpLysLysAsnGlyTyrTrpLysSerArgAlaGlyValIleAsn  240
             |||||||||||||||||||||||||||||||||||||
Db   425 TGGGTTCAAGGTGGAAGAAAATGGGTGGAACAAGTGCAGGGAAGAGAGTGATCAAC  484

Qy   241 LysGluAspPheValAlaLeuGluArgLeuThrGlnArgGlyMetAspIleGlnTrpMetHis  260
             |||||||||||||||||||||||||||||||||||||
Db   485 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT  544

Qy   261 ValProGlyHisSerGlyPheIleGlyAsnGluGlyAlaAspArgLeuAlaAsgArgGluGly  280
             |||||||||||||||||||||||||||||||||||||
Db   545 GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA  604

Qy   281 AlaLysGlnSerGluAsp  286
             ||||||||||
Db   605 GCTAAACAATCGGAAGAC  622
```

RESULT 6

```
ADA05717
ID   ADA05717 standard; cDNA; 716 BP.
XX
AC   ADA05717;
XX
DT   06-NOV-2003  (first entry)
XX
DE   Human NOV15a encoding cDNA SEQ ID NO:77.
XX
KW   human; NOVX; antidiabetic; anorectic; antibacterial; virucide;
KW   immunomodulator; cytostatic; nootropic; neuroprotective;
KW   antiparkinsonian; antilipaemic; gene therapy; human disease;
KW   metabolic disorder; diabetes; obesity; infection; cachexia; cancer;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   CDS            10..621
```

```
FT                            /*tag= a
FT                            /product= "NOV15a"
XX
PN   WO2003029424-A2.
XX
PD   10-APR-2003.
XX
PF   02-OCT-2002;  2002WO-US031373.
XX
PR   02-OCT-2001;  2001US-0326483P.
PR   05-OCT-2001;  2001US-0327435P.
PR   05-OCT-2001;  2001US-0327449P.
PR   09-OCT-2001;  2001US-0327917P.
PR   09-OCT-2001;  2001US-0328029P.
PR   09-OCT-2001;  2001US-0328044P.
PR   09-OCT-2001;  2001US-0328056P.
PR   12-OCT-2001;  2001US-0328849P.
PR   15-OCT-2001;  2001US-0329414P.
PR   17-OCT-2001;  2001US-0330142P.
PR   18-OCT-2001;  2001US-0330309P.
PR   22-OCT-2001;  2001US-0341058P.
PR   24-OCT-2001;  2001US-0339266P.
PR   24-OCT-2001;  2001US-0343629P.
PR   29-OCT-2001;  2001US-0345575P.
PR   01-NOV-2001;  2001US-0346357P.
PR   17-APR-2002;  2002US-0373260P.
PR   19-APR-2002;  2002US-0373815P.
PR   19-APR-2002;  2002US-0373817P.
PR   19-APR-2002;  2002US-0373826P.
PR   19-APR-2002;  2002US-0373884P.
PR   22-APR-2002;  2002US-0374977P.
PR   16-MAY-2002;  2002US-0381037P.
PR   16-MAY-2002;  2002US-0381038P.
PR   16-MAY-2002;  2002US-0381042P.
PR   17-MAY-2002;  2002US-0381642P.
PR   28-MAY-2002;  2002US-0383656P.
PR   29-MAY-2002;  2002US-0383831P.
PR   25-JUN-2002;  2002US-0391335P.
PR   01-OCT-2002;  2002US-00262511.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Smithson G,  Millet I,  Peyman JA,  Kekuda R,  Ju J,  Li L,  Guo X;
PI   Patturajan M,  Spytek KA,  Edinger SR,  Ellerman K,  Malyankar UM;
PI   Ort T,  Gorman L,  Zerhusen BD,  Anderson DW,  Zhong M,  Catterton E;
PI   Ji W,  Miller CE,  Rastelli L,  Stone DJ,  Pena CEA,  Shenoy SG;
PI   Shimkets RA,  Rothenberg ME,  Leach MD,  Agee ML,  Berghs C,  DiPippo VA;
PI   Eisen AJ,  Gangolli EA,  Rieger DK,  Spaderna SK;
XX
DR   WPI; 2003-381625/36.
DR   P-PSDB; ADA05718.
XX
PT   New NOVX polypeptides and nucleic acids, useful for diagnosing,
PT   preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,
PT   cancer or dyslipidemia, and in chromosome mapping, tissue typing or
PT   pharmacogenomics.
XX
PS   Claim 20; Page 159; 586pp; English.
XX
CC   The present invention describes NOVX proteins, where X can be 1 to 55
CC   (e.g. NOV1). Also described: (1) a composition comprising a polypeptide
CC   described above and a carrier; (2) a kit comprising, in one or more
CC   containers, the composition described above; (3) an isolated nucleic acid
CC   molecule which encodes a NOVX protein of the invention; (4) a vector
CC   comprising the nucleic acid molecule described above; (5) a cell
CC   comprising the above vector; (6) an antibody that immunospecifically
CC   binds to the polypeptide described above; (7) methods for determining the
CC   presence or amount of the above polypeptide or nucleic acid molecule in a
CC   sample; (8) methods for determining the presence of or predisposition to
CC   a disease associated with altered levels of expression of the above
CC   polypeptide or nucleic acid molecule in a first mammalian subject; (9) a
CC   method of identifying an agent that binds to the polypeptide described
CC   above; (10) a method for identifying a potential therapeutic agent for
```

CC   use in treating a pathology that is related to an aberrant expression or
CC   aberrant physiological interactions of the polypeptide; (11) a method of
CC   screening for a modulator of activity or of latency or predisposition to
CC   a pathology associated with the polypeptide; (12) a method for modulating
CC   the activity of the polypeptide described above; (13) methods of treating
CC   or preventing a pathology associated with the above polypeptide in a
CC   mammal; and (14) a method for producing the above polypeptide. NOVX
CC   sequences have antidiabetic, anorectic, antibacterial, virucide,
CC   immunomodulator, cytostatic, nootropic, neuroprotective, antiparkinsonian
CC   and antilipaemic activities, and can be used in gene therapy. The
CC   polypeptide is useful in manufacturing a medicament for treating a
CC   syndrome associated with a human disease. The polypeptide or the nucleic
CC   acid molecule may be used to diagnose, treat or prevent metabolic
CC   disorders such as diabetes or obesity, infections, cachexia, cancer,
CC   neurodegenerative disorders such as Alzheimer's disease or Parkinson's
CC   disease, immune disorders, haematopoietic disorders and various
CC   dyslipidaemias. The nucleic acids can also be used as hybridisation
CC   probes, in chromosome mapping, tissue typing, preventive medicine and
CC   pharmacogenomics. The present sequence encodes a human NOVX protein from
CC   the present invention.
XX
SQ   Sequence 716 BP; 185 A; 152 C; 217 G; 162 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   716
Score: 1.54e-83
Percent Similarity: 1045.50
Best Local Similarity: 70.63%
Query Match: 70.28%
DB: 67.63%
                           7

US-10-054-313-1 (1-286) x ADA05717 (1-716)

Qy   1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaAlaLeuAlaAlaLeuProCysArgArg 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10 ATGAGCTGGTTGTTCTGTTCCTGGCCCACAGAGTCGCCTGGCCCTGCCCTGCCGCCGCG 69

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgLysThrGlyValPhe 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   70 GGCTCTCGCGGGGTTCGGGATGTTCATACCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 129

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAlaSerArgArgPheProAlaAlaArgPheLysLys 60
         ||||||||||||||||||||
Db 130 CTGACCTGGAATGAGTGC-------------------------------------- 147

Qy   61 PheAlaAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80

Db 147 -- 147

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100

Db 147 -- 147

Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120

Db 147 -- 147

Qy 121 SerValGluProAlaProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140

Db 148 -------------------AGAGACACGTTTCCTACATGGGAGACTTCGTC 180

Qy 141 ValValTyrThrAspSerMetPheThrIleAsnGlyArgArgLysProArgAlaGlyIle 160
         ||||||||||||||||||||||||||||||||||:::||||||||||||||
Db 181 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTGAAGGCCGCAGCAGGAATC 240

Qy 161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleIleArgLeuProGlyArgGln 180
         ||||||||||||||||||||||||||||||||||||||||||||
Db 241 GGCGTTTACTGCGGGGCCCGGCCGGCCCATCCTTAAATGTAGGCATTAGACTTCCTGGGCCGCAG 300

Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA 360

201 AsnIleAsnLysLeuValValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
Qy
Db 361 AACATCAATAAACTGGTTCTGTATACAGACAGCATGTTTACGATAAATGGTATAACTAAC 420

221 TrpValGlnGlyTrpLysLysAsnGlyTrpSerThrSerAlaGlyLysLysGluValIleAsn 240
Qy
Db 421 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAGCACAAGTCGCAGGGAAAGAGGTGATCAAC 480

241 LysGluAspPheValAlaLeuGluGlyArgLeuThrGlyMetAspIleTrpMetHis 260
Qy
Db 481 AAAGAGGACTTTGTGGCACTGGAGGGGCGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 540

261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyGly 280
Qy
Db 541 GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA 600

281 AlaLysGlnSerGluAsp 286
Qy
Db 601 GCTAAACAATCGGAAGAC 618

RESULT 7
ADA05721
ID   ADA05721 standard; cDNA; 586 BP.
XX
AC   ADA05721;
XX
DT   06-NOV-2003 (first entry)
XX
DE   Human NOV15c encoding cDNA SEQ ID NO:81.
XX
KW   human; NOVX; antidiabetic; anorectic; antibacterial; virucide;
KW   immunomodulator; cytostatic; nootropic; neuroprotective;
KW   antiparkinsonian; antilipaemic; gene therapy; human disease;
KW   metabolic disorder; diabetes; obesity; infection; cachexia; cancer;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   2..586
FT   /*tag= a
FT   /product= "NOV15c"
XX
PN   WO2003029424-A2.
XX
PD   10-APR-2003.
XX
PF   02-OCT-2002; 2002WO-US031373.
XX
PR   02-OCT-2002; 2001US-0326483P.
PR   05-OCT-2001; 2001US-0327435P.
PR   05-OCT-2001; 2001US-0327449P.
PR   09-OCT-2001; 2001US-0327917P.
PR   09-OCT-2001; 2001US-0328029P.
PR   09-OCT-2001; 2001US-0328044P.
PR   09-OCT-2001; 2001US-0328056P.
PR   12-OCT-2001; 2001US-0328849P.
PR   15-OCT-2001; 2001US-0329414P.
PR   17-OCT-2001; 2001US-0330142P.
PR   18-OCT-2001; 2001US-0330309P.
PR   22-OCT-2001; 2001US-0341058P.
PR   24-OCT-2001; 2001US-0339266P.
PR   24-OCT-2001; 2001US-0343629P.
PR   29-OCT-2001; 2001US-0349575P.
PR   01-NOV-2001; 2001US-0363357P.
PR   17-APR-2002; 2002US-0373260P.
PR   19-APR-2002; 2002US-0373815P.
PR   19-APR-2002; 2002US-0373817P.
PR   19-APR-2002; 2002US-0373826P.
PR   19-APR-2002; 2002US-0373884P.
PR   22-APR-2002; 2002US-0374977P.
PR   16-MAY-2002; 2002US-0381037P.

```
PR    16-MAY-2002;  2002US-0381038P.
PR    16-MAY-2002;  2002US-0381042P.
PR    17-MAY-2002;  2002US-0381642P.
PR    28-MAY-2002;  2002US-0383556P.
PR    29-MAY-2002;  2002US-0383831P.
PR    25-JUN-2002;  2002US-0391335P.
PR    01-OCT-2002;  2002US-0262511.
XX
PA    (CURA-) CURAGEN CORP.
XX
PI    Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;
PI    Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;
PI    Ort T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Catterton E;
PI    Ji W, Miller CE, Rastelli L, Stone DJ, Pena CEA, Shenoy SG;
PI    Shimkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dippipo VA;
PI    Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;
XX
DR    WPI; 2003-381626/36.
DR    P-PSDB; ADA05722.
XX
PT    New NOVX polypeptides and nucleic acids, useful for diagnosing,
PT    preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,
PT    cancer or dyslipidemia, and in chromosome mapping, tissue typing or
PT    pharmacogenomics.
XX
PS    Claim 20; Page 159-160; 586pp; English.
XX
CC    The present invention describes NOVX proteins, where X can be 1 to 55
CC    (e.g. NOV1). Also described: (1) a composition comprising a polypeptide
CC    described above and a carrier; (2) a kit comprising, in one or more
CC    containers, the composition described above; (3) an isolated nucleic acid
CC    molecule which encodes a NOVX protein of the invention; (4) a vector
CC    comprising the nucleic acid molecule described above; (5) a cell
CC    comprising the above vector; (6) an antibody that immunospecifically
CC    binds to the polypeptide described above; (7) methods for determining the
CC    presence or amount of the above polypeptide or nucleic acid molecule in a
CC    sample; (8) methods for determining the presence of or predisposition to
CC    a disease associated with altered levels of expression of the above
CC    polypeptide or nucleic acid molecule in a first mammalian subject; (9) a
CC    method of identifying an agent that binds to the polypeptide described
CC    above; (10) a method for identifying a potential therapeutic agent for
CC    use in treating a pathology that is related to an aberrant expression or
CC    aberrant physiological interactions of the polypeptide; (11) a method of
CC    screening for a modulator of activity or of latency or predisposition to
CC    a pathology associated with the polypeptide; (12) a method for modulating
CC    the activity of the polypeptide described above; (13) methods of treating
CC    or preventing a pathology associated with the above polypeptide in a
CC    mammal; and (14) a method for producing the above polypeptide. NOVX
CC    sequences have antidiabetic, anorectic, antibacterial, virucide,
CC    immunomodulator, cytostatic, nootropic, neuroprotective, antiparkinsonian
CC    and antilipaemic activities, and can be used in gene therapy. The
CC    polypeptide is useful in manufacturing a medicament for treating a
CC    syndrome associated with a human disease. The polypeptide or the nucleic
CC    acid molecule may be used to diagnose, treat or prevent metabolic
CC    disorders such as diabetes or obesity, infections, cachexia, cancer,
CC    neurodegenerative disorders such as Alzheimer's disease or Parkinson's
CC    disease, immune disorders, haematopoietic disorders and various
CC    dyslipidaemias. The nucleic acids can also be used as hybridisation
CC    probes, in chromosome mapping, tissue typing, preventive medicine and
CC    pharmacogenomics. The present sequence encodes a human NOVX protein from
CC    the present invention.
XX
SQ    Sequence 586 BP; 161 A; 125 C; 177 G; 123 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score:                    979.50
Percent Similarity:       69.60%        Length:           586
Best Local Similarity:    68.86%        Matches:          188
Query Match:              63.36%        Conservative:       2
DB:                                     Mismatches:         0
                                        Indels:            83
                          8.5e-78       Gaps:               1

US-10-054-313-1 (1-286) x ADA05721 (1-586)
```

```
Qy     14  AlaAlaAlaLeuProCysArgArgArgGlySerArgArgGlyPheGlyMetPheTyrAlaValArgArg  33
              :::  ||:||||  ||||   |  |   |||||||||||||||||||||||
Db      8  TCGGCCTTGCCCTGCCGCGCCGGCTCCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGG  67

Qy     34  GlyArgGlyThrGlyValPheLeuThrTrpAsnGluCysArgAlaGlnValAspArgPhe  53
              ||:||||||||||||||||||||||||||||||||||||||||||||||
Db     68  GGCCCAAGACCGGGGTCTTTCTGACCTGGAATGAGTGC--------------------  106

Qy     54  ProAlaAlaArgPheLysPheAlaThrGluAspGluAlaTrpAlaPheValArgLys  73
              -------------------------------------------------------
Db    106  --------------------------------------------------------  106

Qy     74  SerAlaSerProGluValSerGluGlyHisGluAsnGlnHisGlyGlnGluSerGluAla  93
              ---------------------------------------------------------
Db    106  --------------------------------------------------------  106

Qy     94  LysProGlyLysArgGluProLeuArgGlyHisGluSerAlaGlnPro  113
              --------------------------------------------------
Db    106  ------------------------------------------------  106

Qy    114  TyrAlaLysHisMetLysProSerValGluProAlaProValSerArgArgAspThrPhe  133
                                                        ||:||
Db    107  ---------------------------------------------AGAGACACGTTT  118

Qy    134  SerTyrMetGlyAspPheValValTyrThrAspGlyCysCysSerSerAsnGlyArg  153
              ||||:||:||:|||||:|:|  |:|| |:|  ||| ||||||||  ||||:||
Db    119  TCCTACATGGGAGACTTCGTCGTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGT  178

Qy    154  ArgLysProArgAlaGlyIleGlyValTyrIleGlyProGlyHisProLeuAsnValGly  173
              |:||||  |  ||||||  |||||  ||| ||:||||||||  ||:|| ||| |:|
Db    179  AGAAGGCCGCGAGCAGGAATCGGCGTTTACTGGGGGCCGGGCCATCCTTTAAATGTAGGC  238

Qy    174  IleArgLeuProGlyArgGlnThrArgGlnArgAlaGlnIleHisAlaAlaCysLysAla  193
              |  ||:  |||  |||||  |  ||| ||  |||||||||  ||:||| ||||||
Db    239  ATTAGACTTCCTGGCCGGCAGACAACAAAGAGCGGAAATTCATGCAGCCTGCAAAGCC  298

Qy    194  IleGluGlnAlaLysThrGlnAsnIleAsnLysValLeuValLeuTyrThrAspSerMetPhe  213
              |  |||||||:  |:|||  |||||| || |:|| ||: ||||||| ||:|
Db    299  ATTGAACAAGCAAGACTCAAAACATCAAATAAACTGTTCTGTATACAGACAGTATGTTT  358

Qy    214  ThrIleAsnGlyIleThrArgAsnTrpValGlnGlyTrpLysAsnGlyTrpLysThrSer  233
              || |||| ||:||:  |  ||||||| |:|||  ||||||  |||| |:|||
Db    359  ACGATAAATGGTATAACTAACTGGGTTCAAGGTTGGAAGAAAAATGGGTGAAGACAAGT  418

Qy    234  AlaGlyLysGluValIleLeuAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGly  253
              |||||  |:|||:|  ||||||| || |||||||| ||||| |:||||||||||||
Db    419  GCAGGGAAAGAGGTGATCAACAAGAGACTTTGTGCACTGGAGAGGCTTACCCAGGGG  478

Qy    254  MetAspPheGlnTrpMetHisValProGlyHisSerGlyPheIleGlyPheLeuAsnGluAla  273
              |||||||||:|| |||| |||| ||:  |:|||:||||: ||||||:|
Db    479  ATGGACATTCAGTGGATGCATGTTCCTGGTCATTCGGGATTTATAGGCAATCGAAGAAGCT  538

Qy    274  AspArgLeuAlaArgGluGluValGlnAlaLysGlnSerGlnGluAsp  286
              |||||| |  |||||||| |:|| |||||| |:|| |:|||||
Db    539  GACAGATTAGCCAGAGAAGGAGCTAAACAATCGGAAGAC  577
```

```
RESULT 8
ABV45787
ID     ABV45787  standard;  cDNA;  545 BP.
XX
AC     ABV45787;
XX
DT     16-SEP-2002  (first entry)
XX
DE     Human prostate expression marker cDNA 45778.
XX
KW     Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW     pharmacogenomic marker; gene; ss.
XX
OS     Homo sapiens.
XX
PN     WO200160860-A2.
```

```
Qy    133  PheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGly  152
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    364  TTTTCCTACATGGGAGACTTCGTCGTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGG  423

Qy    153  ArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyHisProLeuAsnVal  172
           |||||:::|||||||||||||||||||||||||||||||:::::||||:::|||
Db    424  CGTAGAAGGCCCGAGCAGGAATCGGCGTTTACTGGGGCCCATCCTTTAAATGTA  483

Qy    173  GlyIleArgLeuProGlyArgGlnThrArgAsnGlnAsnGlnAsnGlnGlyLys  192
           |||||||:::|||||||||||||||||||||||||||||||||||||||||||
Db    484  GGCATTAGACTTCCTGGGCGGCAGACAACCAAGAGCGGAAATTCATGCAACCTGCAA  543
```

RESULT 9
ADA05723

ID   ADA05723 standard; cDNA; 457 BP.
XX
AC   ADA05723;
XX
DT   06-NOV-2003 (first entry)
XX
DE   Human NOV15d encoding cDNA SEQ ID NO:83.
XX
KW   human; NOVX; antidiabetic; anorectic; antibacterial; virucide;
KW   immunomodulator; cytostatic; nootropic; neuroprotective;
KW   antiparkinsonian; antilipaemic; gene therapy; human disease;
KW   metabolic disorder; diabetes; obesity; infection; cachexia; cancer;
KW   neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW   immune disorder; haematopoietic disorder; dyslipidaemia; gene; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   2..457
FT   /*tag= a
FT   /product= "NOV15d"
XX
PN   WO2003029424-A2.
XX
PD   10-APR-2003.
XX
PF   02-OCT-2002; 2002WO-US031373.
XX
PR   02-OCT-2001; 2001US-0326483P.
PR   05-OCT-2001; 2001US-0327435P.
PR   05-OCT-2001; 2001US-0327449P.
PR   09-OCT-2001; 2001US-0327917P.
PR   09-OCT-2001; 2001US-0328029P.
PR   09-OCT-2001; 2001US-0328044P.
PR   09-OCT-2001; 2001US-0328056P.
PR   12-OCT-2001; 2001US-0328849P.
PR   15-OCT-2001; 2001US-0329414P.
PR   17-OCT-2001; 2001US-0330142P.
PR   18-OCT-2001; 2001US-0330309P.
PR   22-OCT-2001; 2001US-0341058P.
PR   24-OCT-2001; 2001US-0339266P.
PR   24-OCT-2001; 2001US-0343629P.
PR   29-OCT-2001; 2001US-0349575P.
PR   01-NOV-2001; 2001US-0363357P.
PR   17-APR-2002; 2002US-0373260P.
PR   19-APR-2002; 2002US-0373815P.
PR   19-APR-2002; 2002US-0373817P.
PR   19-APR-2002; 2002US-0373826P.
PR   19-APR-2002; 2002US-0373884P.
PR   22-APR-2002; 2002US-0374977P.
PR   16-MAY-2002; 2002US-0381037P.
PR   16-MAY-2002; 2002US-0381042P.
PR   17-MAY-2002; 2002US-0381642P.
PR   28-MAY-2002; 2002US-0383656P.
PR   29-MAY-2002; 2002US-0393831P.
PR   25-JUN-2002; 2002US-0391335P.
PR   01-OCT-2002; 2002US-00262511.
XX
```

```
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001;  2001WO-US005171.
XX
PR   17-FEB-2000;  2000US-0183319P.
PR   16-MAR-2000;  2000US-0189862P.
PR   25-MAY-2000;  2000US-0207454P.
PR   09-JUN-2000;  2000US-0211314P.
PR   18-JUL-2000;  2000US-0219007P.
PR   13-DEC-2000;  2000US-0255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS   Claim 1; Page 9051; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) or its
CC   specification or its complement. (I) is useful for: (a) assessing whether
CC   a patient is afflicted with prostate cancer; (b) monitoring the
CC   progression of prostate cancer in a patient; (c) assessing the efficacy
CC   of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC   the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC   determining whether prostate cancer has metastasized in a patient; (h)
CC   assessing the aggressiveness or indolence of prostate cancer in a patient
CC   ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ   Sequence 545 BP; 136 A; 129 C; 178 G; 102 T; 0 U; 0 Other;
```

Alignment Scores:
```
Pred. No.:               3.15e-71
Score:                   905.00
Percent Similarity:      94.44%
Best Local Similarity:   92.78%
Query Match:             58.54%
DB:                      5
```
```
Length:         545
Matches:        167
Conservative:     3
Mismatches:      10
Indels:           0
Gaps:             0
```

US-10-054-313-1 (1-286) x ABV45787 (1-545)

```
Qy    13   LeuAlaAlaLeuProCysArgArgGlySerArgGlyPheGlyMetPheTyrAlaValArg  32
           |:::                          |||||||||||||||||||||||||||||
Db    4    ATTGGAGCTCCACCGCGGTGGCGGCCCAGGTCGCGGGTTCGGGATGTTCTATGCCGTGAGG  63

Qy    33   ArgGlyArgLysThrGlyValPheLeuThrTrpAsnGluCysArgArgAlaGlnValAspArg  52
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    64   AGGGGCCCGACACTGGGGTCTTCTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGG  123

Qy    53   PheProAlaAlaArgPheLysLysPheAlaThrGluAspTrpAlaTrpAlaPheValArg  72
           |||                          |||||||||||||||||||||||||||||||
Db    124  TTTCCTGCTGCCAGATTTAAGAAGTTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGG  183

Qy    73   LysSerAlaSerProGluValSerGluGluValHisGluAsnGlnHisGlyGlnGlySerGlu  92
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    184  AAATCTGCAAGCCCGAAGTTTCAGAGGGCATGAAAATCAACATGGACAAGAATCGGAG  243

Qy    93   AlaLysProGlyLysArgLeuArgGluProLeuAspGlyAlaAspGlyHisGluSerAlaGln  112
           ||||||||||||||||||||||||||||||||||||||||||||||||:::
Db    244  GCGAAAGCCAGCAAGCGACTCCGTGAGCCACTGGATGGAGCATGAGATGGAGCATGAAAGCGCAGAG  303

Qy    113  ProTyrAlaLysHisMetLysProSerValGluProAlaProProValSerArgAspThr  132
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    304  CCGTATGCAAAGCACATGAAAGCCAGCGTGGAGCCCGCGCCTCCCAGTTAGCAGAGACACG  363
```

PA (CURA-) CURAGEN CORP.
XX
PI Smithson G, Millet I, Peyman JA, Kekuda R, Ju J, Li L, Guo X;
PI Patturajan M, Spytek KA, Edinger SR, Ellerman K, Malyankar UM;
PI Ort T, Gorman L, Zerhusen BD, Anderson DW, Zhong M, Catterton B;
PI Ji W, Miller CE, Rastelli L, Stone DJ, Pena CEA, Shenoy SG;
PI Shimkets RA, Rothenberg ME, Leach MD, Agee ML, Berghs C, Dipippo VA;
PI Eisen AJ, Gangolli EA, Rieger DK, Spaderna SK;
XX
DR WPI; 2003-381626/36.
DR P-PSDB; ADA05724.
XX
PT New NOVX polypeptides and nucleic acids, useful for diagnosing,
PT preventing or treating NOVX-associated disorders, e.g. diabetes, obesity,
PT cancer or dyslipidemia, and in chromosome mapping, tissue typing or
PT pharmacogenomics.
XX
PS Claim 20; Page 160; 586pp; English.
XX
CC The present invention describes NOVX proteins, where X can be 1 to 55
CC (e.g. NOV1). Also described: (1) a composition comprising a polypeptide
CC described above and a carrier; (2) a kit comprising, in one or more
CC containers, the composition described above; (3) an isolated nucleic acid
CC molecule which encodes a NOVX protein of the invention; (4) a vector
CC comprising the nucleic acid molecule described above; (5) a cell
CC comprising the above vector; (6) an antibody that immunospecifically
CC binds to the polypeptide described above; (7) methods for determining the
CC presence or amount of the above polypeptide or nucleic acid molecule in a
CC sample; (8) methods for determining the presence of or predisposition to
CC a disease associated with altered levels of expression of the above
CC polypeptide or nucleic acid molecule in a first mammalian subject; (9) a
CC method of identifying an agent that binds to the polypeptide described
CC above; (10) a method for identifying a potential therapeutic agent for
CC use in treating a pathology that is related to an aberrant expression or
CC aberrant physiological interactions of the polypeptide; (11) a method of
CC screening for a modulator of activity of or latency or predisposition to
CC a pathology associated with the polypeptide; (12) a method for modulating
CC the activity of the polypeptide described above; (13) methods of treating
CC or preventing a pathology associated with the above polypeptide in a
CC mammal; and (14) a method for producing the above polypeptide. NOVX
CC sequences have antidiabetic, anorectic, antibacterial, virucide,
CC immunomodulator, cytostatic, nootropic, neuroprotective, antiparkinsonian
CC and antilipaemic activities, and can be used in gene therapy. The
CC polypeptide is useful in manufacturing a medicament for treating a
CC syndrome associated with a human disease. The polypeptide or the nucleic
CC acid molecule may be used to diagnose, treat or prevent metabolic
CC disorders such as diabetes or obesity, infections, cachexia, cancer,
CC neurodegenerative disorders such as Alzheimer's disease or Parkinson's
CC disease, immune disorders, haematopoietic disorders and various
CC dyslipidaemias. The nucleic acids can also be used as hybridisation
CC probes, in chromosome mapping, tissue typing, preventive medicine and
CC pharmacogenomics. The present sequence encodes a human NOVX protein from
CC the present invention.
XX
SQ Sequence 457 BP; 139 A; 89 C; 134 G; 95 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:
Score: 2.64e-61
Percent Similarity: 792.00   Length: 457
Best Local Similarity: 100.00%   Matches: 145
Query Match: 99.32%   Conservative: 1
DB: 51.23%   Mismatches: 0
                                      Indels: 0
                        7   Gaps: 0

US-10-054-313-1 (1-286) x ADA05723 (1-457)

Qy 137 GlyAspPheValValTyrThrArgGlyCysCysSerAsnGlyArgArgPro 156
       |||||||||||||||||||||||||||||||||||||||||||||:::||
Db 11 GGAGACTTCGTCGTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCG 70

Qy 157 ArgAlaGlyIleGlyValTyrTrpGlyHisProLeuAsnValGlyIleArgLeu 176
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 71 CGAGCAGGAATCGGCGTTACTGGGGGGCCGGGCCATCCTTTAAATGTAGGCATTAGACTT 130

Qy 177 ProGlyArgGlnThrAsnGlnArgAlaHisAlaAlaAlaCysLysAlaIleGluGlnGln 196
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 131 CCTGGCGGCAGACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAGACCCATTGAACAA 190

Qy 197 AlaLysThrGlnAsnIleAsnLysValLeuTyrThrAspSerMetPheThrIleAsn 216
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 191 GCAAAGACTCAAACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAAT 250

Qy 217 GlyIleThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLys 236
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 251 GGTATAACTAACTGGGTTCAAGGTTGGAAGAAAAATGGTTGGAAGACAAGTCAGGGAAA 310

Qy 237 GluValIleAsnLysGluAspPheValAlaLeuGluThrGlnGlyMetAspIle 256
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 311 GAGGTGATCAACAAAGAGGACTTTGTGGCACTGGAGACGCAAGGCTTACCCAGGGATGGACATT 370

Qy 257 GlnTrpMetHisValProGlyHisSerGlyPheIleGlyLysAsnGluGluLysArgLeu 276
       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db 371 CAGTGGATGCATGTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTA 430

Qy 277 AlaArgGluGlyAlaLys 282
       |||||||||||
Db 431 GCCAGAGAGGAGCTAAA 448

RESULT 10
ABV15988
ID ABV15988 standard; cDNA; 691 BP.
XX
AC ABV15988;
XX
DT 13-SEP-2002 (first entry)
XX
DE Human prostate expression marker cDNA 15979.
XX
KW Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW pharmacogenomic marker; gene; ss.
XX
OS Homo sapiens.
XX
PN WO20016C0860-A2.
XX
PD 23-AUG-2001.
XX
PF 20-FEB-2001; 2001WO-US005171.
XX
PR 17-FEB-2000; 2000US-0183319P.
PR 16-MAR-2000; 2000US-0189862P.
PR 25-MAY-2000; 2000US-0207454P.
PR 09-JUN-2000; 2000US-0211314P.
PR 18-JUL-2000; 2000US-0219007P.
PR 13-DEC-2000; 2000US-0255281P.
XX
PA (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI Schlegel R, Endege WO, Monahan JE;
XX
DR WPI; 2001-662795/76.
XX
PT Novel isolated nucleic acid molecule associated with cancerous state of
PT prostate cells and correlating with presence of prostate cancer, useful
PT for detecting presence of prostate cancer, stage of prostate cancer.
XX
PS Claim 1; Page 2678; 11750pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule (I) comprising
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC specification or its complement. (I) is useful for: (a) assessing whether
CC a patient is afflicted with prostate cancer; (b) monitoring the
CC progression of prostate cancer in a patient; (c) assessing the efficacy
CC of a test compound to inhibit prostate cancer in a patient; (d) assessing
CC the efficacy of a therapy for inhibiting prostate cancer in a patient;
CC (e) selecting a composition for inhibiting prostate cancer in a patient;

CC   (f) assessing the prostate cell carcinogenic potential of a compound; (g)
CC   determining whether prostate cancer has metastasized in a patient; (h)
CC   assessing the aggressiveness or indolence of prostate cancer in a patient
CC ; (I) is also useful as a pharmacodyanamic or pharmacogenomic marker
XX
SQ   Sequence 691 BP; 186 A; 155 C; 199 G; 144 T; 0 U; 7 Other;

Alignment Scores:
Pred. No.:   5.86e-50   Length: 691
Score: 666.50   Matches: 159
Percent Similarity: 73.82%   Conservative: 13
Best Local Similarity: 68.24%   Mismatches: 34
Query Match: 43.11%   Indels: 28
DB: Gaps: 8

US-10-054-313-1 (1-286) x ABV15988 (1-691)

Qy   24 GlyPheGlyMetPheTyrAlaValAlaArgGlyArgGlyValPheLeuThrTrp 43
               ||||||||||||||||||||||||||||||||||||||||
Db   3 GGGTTCGGGATGTTCTATGCCGTGGCAGGGCCGCAAGACTGGGGTCTTTCTGACCTGG 62

Qy   44 AsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLysPheAlaThr 63
               |||||||||||||||||||||||||||||||||||||||:::
Db   63 AATGAGTGCAGAGCACAGGTGCGACCCGTTTCCTGCTGCCAGATTTAAGAAGTTTGCCACA 122

Qy   64 GluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHis 83
               ||||||||||||||||||||||||||||||||||||||||
Db   123 GAGGATGAGGCCTGGGCTTTCGTCAGGAAATCTGCAAGCCGGAGTTTCAGAGGGGCAT 182

Qy   84 GluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgGluArgGluProLeu 103
               ||||||||||||||||||||||||||||||||||||
Db   183 GAAAATCAACATGGACAAGAATCGGAGGCGAAAGCCAGCAGGCGACTCCGTGAGCCACTG 242

Qy   104 AspGlyAspGlyHisGluSerAlaGln-ProTyrAlaLysHis-MetLysProSerVal- 122
               |||||||||||||||||||||||||| ||||||||||| |||||||||||||
Db   243 GATGGAGATGGACATGAAAGCGCAGAGCCCGTATGCCAAAGCCACCATGAAGCCCGAGCCGT 302

Qy   123 --GluProAlaProValSerArgAspThr-PheSerTyrMet-GlyAspPheVal-V 141
                 ||||||||||||||||||||||||| ||||||||||| |||||
Db   303 GGAGCCCCGGCGCCCTCCAGTTAGCAGCGAGCACCGTTTTCCTACATGGGGAGACTTCGTTCG 362

Qy   141 alValTyrThrAspGly-CysCysSerSerAsnGlyArgArgLys------ProArgAla 158
               |||||||||||||| ||||||||||||||||||| ||||||||
Db   363 TCGTCTACACTGATGGTTGCTGCTCCAGTAATGGGCGTANGAAGCCCCGAGGCAG 422

Qy   159 GlyIleGlyValTyrTrp---GlyProGlyHisProLeuAsn--ValGlyIleArgLeu 176
               |||||||||||||||| |||||||||||||||||| ||||||||||||
Db   423 GGAATCGGCGTTTACTGGGGGGCCCCGGCCCATCCTTTAAATGTAGGCCATTAAGGACTT 482

Qy   177 Pro------GlyArg--GlnThrArgAsnGlnArgAlaGluIleHisAlaAla------CysLy 192
               ||| ||||| ||| ||| ||| ||| |||
Db   483 CCTTGGCCGGCGAGCAAACCCAAAGGAAGGCGGGNAAAATTTATGGCCANGCCCTTGCAA 542

Qy   192 sAlaIleGluGlnAlaLysThrGlnAsnIleAsnLys----------------------- 204
               :|:| ||| ||||||
Db   543 AAGCCCATTTGAAACCAAAGGACTTCAAAAAAACCATTCAAATAAAAAACCTGGG 602

Qy   205 -LeuValLeuTyrThrAspSerMetPheThrIle-------AsnGlyIleThrAsnTr 221
               :||||||||| ||||| ||| :::
Db   603 TTTCTTGTTATTCCAAGACAA-GTTNTTGTTTACCGAATTAAAATGGGNTNTTAAACTT 661

Qy   221 pValGlnGlyTrpLysLysAsnGly 229
               |||||||||||:|||
Db   662 AACCTTGGTTTCAAANGGTTNGGG 686

RESULT 11
ADD34573
ID   ADD34573 standard; DNA; 519 BP.
XX
AC   ADD34573;
XX
DT   15-JAN-2004 (first entry)
XX

DE   Mouse mitochondrial DNA sequence SEQ ID NO:2351.
XX
KW   ds; mouse; array; mitochondrial; hybridisation; energy-metabolism;
KW   mitochondrial disease; oxidative phosphorylation dysfunction;
KW   oxidative stress; apoptosis; aging.
XX
OS   Mus musculus.
XX
PN   WO2003020220-A2.
XX
PD   13-MAR-2003.
XX
PF   30-AUG-2002; 2002WO-US027886.
XX
PR   30-AUG-2001; 2001US-0316323P.
PR   31-AUG-2001; 2001CA-0256540.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Wallace DC, Levy S, Kerstann K, Procaccio V;
XX
DR   WPI; 2003-300821/29.
XX
PT   Array containing probes for genes involved in mitochondrial biology,
PT   useful for determining mitochondrial biology gene expression profiles for
PT   use in diagnosing pathologies and identifying biochemical pathways.
XX
PS   Claim 2; SEQ ID NO 2351; 201pp; English.
XX
CC   The invention relates to a novel array comprising at least two isolated
CC   nucleotide molecules, each molecule having a sequence capable of uniquely
CC   hybridising to a nucleic acid molecule which is an expression product of
CC   a gene involved in mitochondrial biology. The array comprises two or more
CC   isolated nucleic acid molecules or spots, each molecule having a sequence
CC   chosen from sequence of 994 human probes and 2046 mouse probes. An array
CC   of the invention is useful for determining an expression profile of a
CC   mouse or human sample containing nucleic acid, by contacting the array
CC   with the sample under conditions allowing selective hybridisation, and
CC   measuring hybridisation of nucleic acid in the sample to the array to
CC   produce an expression profile. The array is also useful for determining
CC   an expression profile of a first labelled sample containing nucleic acid
CC   relative to a second, differently labelled sample containing nucleic
CC   acid. The second sample is a reference or a standard. An array is useful
CC   for determining an expression profile diagnostic of an energy-metabolism-
CC   related physiological condition. An array of the invention is useful for
CC   determining mitochondrial biology gene expression profiles of organisms,
CC   such as human, mice and closely-related species, tissue and organs of
CC   such organisms, which are useful for determining expression profiles
CC   diagnostic of energy metabolism-related physiological conditions,
CC   diagnosing such physiological conditions, identifying biochemical
CC   pathways, genes, and mutations involved in such physiological conditions,
CC   identifying therapeutic agents useful for preventing and/or treating such
CC   physiological conditions, evaluating and/or monitoring the efficacy of
CC   such therapies, and creating and identifying animal models of human
CC   energy metabolism-related physiological conditions. An array is also
CC   useful for defining expression signatures or profiles for mitochondrial
CC   diseases, as well as distinguishing clinical disorders that result from
CC   oxidative phosphorylation (OXPHOS) dysfunction, oxidative stress,
CC   apoptosis and aging. An array of the invention contains probes of genes
CC   not previously recognised to participate in mitochondrial biology. The
CC   sequences shown in ADD33224-ADD35260 represent murine mitochondrial DNA
CC   clones used to make the probes of the invention. Some sequences are not
CC   present, these are SEQ ID NO's 295, 1174, 1213, 1700, 1728, 1730, 1905,
CC   1906, 2408 and 2643.
XX
SQ   Sequence 519 BP; 137 A; 120 C; 163 G; 99 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:   9.51e-45   Length: 519
Score: 606.00   Matches: 112
Percent Similarity: 81.46%   Conservative: 11
Best Local Similarity: 74.17%   Mismatches: 7
Query Match: 39.20%   Indels: 21

```
DB:                                      Gaps:      1
US-10-054-313-1 (1-286) x ADD34573 (1-519)

Qy   137 GlyAspPheValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysPro 156
             |||::|||||||||||||||||||||||||||||||:|:::|||
Db     1 GGAGAGTCAGTCGTTGTCTACACGGATGCTGTTGCTCCAGTAATGACGGAAGCGGGCA 60

Qy   157 ArgAlaGlyIleGlyValTyrTrpProGlyProGlyHis-ProLeuAsnValGlyIleArgLe 176
             |||||||||||||||||||||||||||| ||
Db    61 CGAGCAGGAATTGGCGTTTACTGGGGCCCAGGCCACCCT---------- 100

Qy   176 uProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGl 196
                              ::|||||||||||
Db   101 ------------------------TCCTGCAAGGCCATCATGCA 120

Qy   196 nAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAs 216
             |||||||||||||||||:::||:||||||||||||||||
Db   121 AGCCAAGGCTCAGAACATCAGCAAGCTGGTTCTGTACACAGACAGTTCACCATCAA 180

Qy   216 nGlyIleThrAsnTrpValGlnGlyTrpLysAsnGlyTyrTrpLysThrSerAlaGlyLy 236
             ||||||||||||||:||:||:::||||
Db   181 TGGGATAACTAACTGGGTTCAGGCTGGAAGAAGAATGCTGGAACAAGTACAGGGAA 240

Qy   236 sGluValIleAsnLysGluAspPheValAlaLeuGluArgThrGlnGlyMetAspIl 256
             |:::|||||||||||| |||:: ||||||||||||||||||||||||||
Db   241 AGATGTGATCAACAAGAGGACTTCATGGAGCTGGACCAGCTCACTCAGGGCATGGACAT 300

Qy   256 eGlnTrpMetHisValProGlyHisSerGlyHisPheIleGlyAsnGluAlaAsnArgLe 276
             ||||||||||::|||||||||::|||||||||||:||
Db   301 CCAGTGGATGCACATTCCTGGTCACTCAGGATTTGTGGGCAATGAAGAGGCCCACAGACT 360

Qy   276 uAlaArgGluGluGlyAlaLysGlnSerGluAsp 286
             |||||||||||||||||||||||||||
Db   361 GGCACGCGAAGGAGCGAAGCAGTCTGAGGAC 391
```

```
RESULT 12
ADD34572/c
ID   ADD34572   standard; DNA; 882 BP.
XX
AC   ADD34572;
XX
DT   15-JAN-2004  (first entry)
XX
DE   Mouse mitochondrial DNA sequence SEQ ID NO:2350.
XX
KW   ds; mouse; array; mitochondrial; hybridisation; energy-metabolism;
KW   mitochondrial disease; oxidative phosphorylation dysfunction;
KW   oxidative stress; apoptosis; aging.
XX
OS   Mus musculus.
XX
PN   WO2003020220-A2.
XX
PD   13-MAR-2003.
XX
PF   30-AUG-2002; 2002WO-US027886.
XX
PR   30-AUG-2001; 2001US-0316323P.
PR   31-AUG-2001; 2001CA-02356540.
XX
PA   (UYEM-) UNIV EMORY.
XX
PI   Wallace DC, Levy S, Kerstann K, Procaccio V;
XX
DR   WPI; 2003-300821/29.
XX
PT   Array containing probes for genes involved in mitochondrial biology,
PT   useful for determining mitochondrial biology gene expression profiles for
PT   use in diagnosing pathologies and identifying biochemical pathways.
XX
PS   Claim 2; SEQ ID NO 2350; 201pp; English.
XX
```

```
CC   The invention relates to a novel array comprising at least two isolated
CC   nucleotide molecules, each molecule having a sequence capable of uniquely
CC   hybridising to a nucleic acid molecule which is an expression product of
CC   a gene involved in mitochondrial biology. The array comprises two or more
CC   isolated nucleic acid molecules or spots, each molecule having a sequence
CC   chosen from sequence of 994 human probes and 2046 mouse probes. An array
CC   of the invention is useful for determining an expression profile of a
CC   mouse or human sample containing nucleic acid, by contacting the array
CC   with the sample under conditions allowing selective hybridisation, and
CC   measuring hybridisation of nucleic acid in the sample to the array to
CC   produce an expression profile. The array is also useful for determining
CC   an expression profile of a first labelled sample containing nucleic acid
CC   relative to a second, differently labelled sample containing nucleic
CC   acid. The second sample is a reference or a standard. An array is useful
CC   for determining an expression profile diagnostic of an energy-metabolism-
CC   related physiological condition. An array of the invention is useful for
CC   determining mitochondrial biology gene expression profiles of organisms,
CC   such as human, mice and closely related species, tissue and organs of
CC   such organisms, which are useful for determining expression profiles
CC   diagnostic of energy metabolism-related physiological conditions
CC   diagnosing such physiological conditions, identifying biochemical
CC   pathways, genes, and mutations involved in such physiological conditions,
CC   identifying therapeutic agents useful for preventing and/or treating such
CC   physiological conditions, evaluating and/or monitoring the efficacy of
CC   such therapies, and creating and identifying animal models of human
CC   energy metabolism-related physiological conditions. An array is also
CC   useful for defining expression signatures or profiles for mitochondrial
CC   diseases, as well as distinguishing clinical disorders that result from
CC   oxidative phosphorylation (OXPHOS) dysfunction, oxidative stress,
CC   apoptosis and aging. An array of the invention contains probes of genes
CC   not previously recognised to participate in mitochondrial biology. The
CC   sequences shown in ADD33224-ADD5260 represent murine mitochondrial DNA
CC   clones used to make the probes of the invention. Some sequences are not
CC   present, these are SEQ ID NO's 295, 1174, 1213, 1700, 1728, 1730, 1905,
CC   1906, 2408 and 2643.
XX
SQ   Sequence 882 BP; 211 A; 231 C; 193 G; 247 T; 0 U; 0 Other;
```

```
Alignment Scores:
Pred. No.:                              Length:              882
Score:                  2.32e-41        Matches:             104
Percent Similarity:     79.17%          Conservative:         10
Best Local Similarity:  72.22%          Mismatches:           10
Query Match:            36.93%          Indels:               20
                                        Gaps:                  1
DB:

US-10-054-313-1 (1-286) x ADD34572 (1-882)

Qy   143 TyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyVal 162
             ||||||||||||||||||||||||||:::::
Db   882 TACACGGATGCTGTTGTCTCCAGTAATGACGGAAGCGGGACCGAGCAGCAGGAATTGGCGTT 823

Qy   163 TyrTrpProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsn 182
             ||:||
Db   822 TACTGGGGCCCAGGCCACCCT--------------------- 802

Qy   183 GlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIle 202
                              :::|||||||||||||||||||||||||
Db   801 ------------------TCCTGCAAGGCCATCATGCAGCCAAGGCTCAGAACATC 763

Qy   203 AsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpVal 222
             :::||||||||||||||||||||||||||||||||||||||||||||||||
Db   762 AGCAAGCTGGTTCTGTACACAGACAGTTCACCATCAATGGGATAACTAACTGGGTT 703

Qy   223 GlnGlyTrpLysAsnGlyTyrTrpLysThrSerAlaGlyLysGluValIleAsnLysGlu 242
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   702 CAGGGCTGGAAGAAGAATGCTGGAACAAGTACAGGGAAAGATGTGATCAACAAGGAG 643

Qy   243 AspPheValAlaLeuGluArgThrGlnGlyMetAspIleGlnTrpMetHisValPro 262
             |||||:::|||||||||||||||||||||||::.:|||
Db   642 GACTTCATGGAGCTGGACCAGCTCACTCAGGGCATGGACATCCAGTGGATGCACATTCCT 583

Qy   263 GlyHisSerGlyHisPheIleGlyAsnGluGluAlaAsnArgLeuAlaArgGluGluAlaGlyAlaLys 282
```

```
Db   582 GGTCACTCAGGATTTGTGGGCAATGAAGAGCCGACAGACTGGCACGGGAAGGAGGCGGAAG 523
         |||||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||

Qy   283 GlnSerGluAsp 286
Db   522 CAGTCTGAGGAC 511


RESULT 13
AAH30241
ID   AAH30241 standard; cDNA; 389 BP.
XX
AC   AAH30241;
XX
DT   27-JUL-2001 (first entry)
XX
DE   Human colon cancer cell line Km12L4-A cDNA library derived sequence #175.
XX
KW   Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;
KW   detection; colon cancer cell line Km12L4-A; ss.
XX
OS   Homo sapiens.
XX
PN   WO200018916-A2.
XX
PD   06-APR-2000.
XX
PF   23-SEP-1999;        99WO-US022226.
XX
PR   28-SEP-1998;        98US-0102161P.
PR   28-SEP-1998;        98US-0102180P.
PR   29-SEP-1998;        98US-0102380P.
PR   08-OCT-1998;        98US-0103815P.
PR   27-OCT-1998;        98US-0105877P.
XX
PA   (CHIR ) CHIRON CORP.
PA   (HYSE-) HYSEQ INC.
XX
PI   Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
PI   Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
PI   Lamson G, Dmananc R, Crkvenjakov R, Dickson M, Dmananc S, Labat I;
PI   Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
XX
DR   WPI; 2000-293155/25.
XX
PT   Polynucleotide library comprising 1079 defined sequences, useful in the
PT   form of an array to detect cancer or susceptibility to cancer.
XX
PS   Claim 1; Page 233; 502pp; English.
XX
CC   The present invention describes a library of polynucleotides comprising
CC   1079 nucleotide sequences (given in AAH30067 to AAH31145). Also described
CC   are: (1) an isolated polynucleotide (I) having at least 90% identity to
CC   one of the 1079 sequences; (2) a recombinant host cell containing (I);
CC   (3) an isolated polypeptide (II) encoded by (I); (4) an antibody that
CC   specifically binds to (II); (5) a vector comprising (I); and (6) a method
CC   of detecting differentially expressed genes correlated with a cancerous
CC   state of a mammalian cell comprising detecting a gene product encoded by
CC   65 of the 1079 sequences given in the specification. The polynucleotides
CC   are used to monitor patients having (or susceptible) to cancer to detect
CC   potentially malignant events at a molecular level before they are
CC   detectable at a gross morphological level. The polynucleotides are also
CC   useful for monitoring the efficacy of various therapies and preventive
CC   interventions. Polynucleotide probes based on the disclosed sequences are
CC   useful for chromosome mapping and detection of transcription levels. The
CC   1079 polynucleotide sequences were derived from a human colon cancer cell
CC   line Km12L4-A cDNA library
XX
SQ   Sequence 389 BP; 78 A; 101 C; 130 G; 80 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:            2.93e-36    Length:          389
Score:               508.50      Matches:         104
Percent Similarity:  56.15%      Conservative:      1
```

Best Local Similarity: 55.61%   Mismatches: 0
Query Match: 32.89%   Indels: 82
DB: 3   Gaps: 1

US-10-054-313-1 (1-286) x AAH30241 (1-389)

```
Qy     1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20
Db    74 ATGAGCTGGCTTCTTCGTTCTGGCCCACAGAGTCGCCCTTGGCCCTGCCCGCCGCGC 133

Qy    21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgValArgGlyLysThrGlyValPhe 40
Db   134 GGTTCTCCGGGGTTCGGATGTTCTATGCCCGTGAGGAGGCCGCAAGAACCGGGGTCTTT 193

Qy    41 LeuThrTrpAsnGluCysArgAlaGlnValAlaAspArgPheProAlaAlaArg-PheLysLy 60
Db   194 CTGACCTGGAATGAGTGCAGAGCACAGGTGGCACCGGTTCCTGCTGCCAGACTTC--- 248

Qy    60 sPheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSe 80
Db   248 ----------------------------------------------------------- 248

Qy    80 rGluGlyHisGluAsnGlnHisGlyGlyGluSerGluAlaLysProGlyLysArgLeuAr 100
Db   248 ----------------------------------------------------------- 248

Qy   100 gGluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaAlaLysHisMetLysPr 120
Db   248 ----------------------------------------------------------- 248

Qy   120 oSerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVa 140
Db   249 ---------------------------------------------------------GT 250

Qy   140 lValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgProArgAlaGlyIl 160
Db   251 CGTCGTCTCACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAGGCCGGCAGGAAT 310

Qy   160 eGlyValTyrTrpGlyTyrProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGl 180
Db   311 CGGCGTTTACTGGGGCCGGGGCCATCCTTAAATGTAGGCATTAGACTTCCTGGGCGCA 370

Qy   180 nThrAsnGlnArgAlaGlu 186
Db   371 GACAAACCAAAGAGCGGAA 389
```

RESULT 14
```
ACH42294
ID   ACH42294 standard; cDNA; 436 BP.
XX
AC   ACH42294;
XX
DT   13-OCT-2003 (first entry)
XX
DE   Human foetal brain cDNA #3661.
XX
KW   Human; ss; sequencing by hybridisation; SBH; expressed sequence tag; EST;
KW   genome mapping; biodiversity; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   US2003073623-A1.
XX
PD   17-APR-2003.
XX
PF   30-JUL-2001; 2001US-00918995.
XX
PR   30-JUL-2001; 2001US-00918995.
XX
PA   (DRMA/) DRMANAC R T.
PA   (LABA/) LABAT I.
PA   (STAC/) STACHE-CRAIN B.
PA   (DICK/) DICKSON M C.
```

PA   (JONE/) JONES L W.
XX
PI   Drmanac RT, Labat I, Stache-Crain B, Dickson MC, Jones LW;
XX
DR   WPI; 2003-615964/58.
XX
PT   New polynucleotide sequences obtained from various cDNA libraries, useful
PT   as hybridization probes, as oligomers for PCR, for chromosome and gene
PT   mapping, in the recombinant production of protein, or in generating
PT   antisense DNA or RNA.
XX
PS   Claim 1; SEQ ID NO 29506; 44pp; English.
XX
CC   The invention relates to an isolated polynucleotide comprising any one of
CC   38043 cDNA sequences, appearing as ACH12789-ACH50831, whose sequence was
CC   determined by the technique of SBH (sequencing by hybridisation). Also
CC   included is a purified polypeptide comprising a sequence corresponding to
CC   a reading frame of the novel polynucleotide. The nucleic acid sequences
CC   are useful in diagnostics as expressed sequence tags (EST) for
CC   identifying expressed genes or for physical mapping of the human genome,
CC   in forensics, in assessing biodiversity, or in identifying mutations
CC   responsible for genetic disorders and other traits. The nucleotide
CC   sequences are also useful as hybridisation probes, as oligomers for PCR,
CC   for chromosome and gene mapping in the recombinant production of
CC   protein, or in generating antisense DNA or RNA. The purified polypeptide
CC   is useful for generating antibodies specific for it. The present sequence
CC   is one of the 38043 isolated cDNA/EST sequences. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from USPTO at
CC   seqdata.uspto.gov/sequence.html?DocID=20030073623
XX
SQ   Sequence 436 BP; 126 A; 79 C; 124 G; 102 T; 0 U; 5 Other;

Alignment Scores:

Pred. No.:	1.31e-32	Length:	436
Score:	468.00	Matches:	86
Percent Similarity:	100.00%	Conservative:	3
Best Local Similarity:	96.63%	Mismatches:	0
Query Match:	30.27%	Indels:	0
DB:	8	Gaps:	0

US-10-054-313-1 (1-286) x ACH42294 (1-436)

```
Qy   198  LysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGly 217
          :::|||||||||||||||||||||||||||||||||||||||||||::::||||||
Db   32   CAGACTCAAAACATCAATAAACTGGTTCTGTATACACAGCAGTATGTTTACTTTAAATGGT 91

Qy   218  IleThrAsnTrpValGlnGlyGlyTrpLysValAsnGlyTrpLysTyrSerAlaGlyLysGlu 237
          :::|||:::|||||||||||||||||||||||||||||||||||||||||||
Db   92   ATTTCTAACTGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGCAAGTGCAGGGAAAGAG 151

Qy   238  ValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnIleMetAspIleGln 257
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   152  GTGATCAACAAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAG 211

Qy   258  TrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAla 277
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   212  TGGATGCATGTTCCTGGTCATTCGGGATTATAGGCAATGAAGAAGTGACAGATTAGCC 271

Qy   278  ArgGluGlyAlaLysGlnSerGluAsp 286
          ||||||||||||||||||||||
Db   272  AGAGAAGGAGCTAAACAATCGGAAGAC 298
```

RESULT 15
ABX41815
ID   ABX41815 standard; cDNA; 473 BP.
XX
AC   ABX41815;
XX
DT   20-FEB-2003 (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #6980.
XX

KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-00960352.
XX
PR   12-JAN-1999;   99US-0115707P.
PR   11-JAN-2000; 2000US-00480902.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
DR   WPI; 2003-110599/10.
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and analysis,
PT   cattle breeding, or for genetically improving cattle.
XX
PS   Claim 2; SEQ ID NO 6980; 245pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived from
CC   cattle, and the LMFD nucleic acid can specifically hybridise to a second
CC   nucleic acid molecule comprising any of 15112 nucleotide sequences,
CC   appearing as ABX34836-ABX49947, or complements of them. Also included are
CC   ; (1) a transformed cell having a nucleic acid comprising an LMFD nucleic
CC   acid linked to a promoter and a 3' non-translated sequence that
CC   functions in the cell to cause termination of transcription and addition
CC   of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC   (2) determining a level or pattern of a molecule in a bovine cell or
CC   tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC   of the 15112 nucleic acid sequences or its complement or fragment) with a
CC   complementary nucleic acid molecule obtained from the bovine cell or
CC   tissue, where hybridisation between the marker nucleic acid and the
CC   complementary nucleic acid permits the detection of the molecule; and (b)
CC   detecting the level or pattern of the complementary nucleic acid, where
CC   the detection of the complementary nucleic acid is predictive of the
CC   level or pattern of the molecule. The LMFD nucleic acid is used for
CC   determining a level or pattern of a molecule in a bovine cell or tissue.
CC   It is useful for genome mapping, gene identification and analysis, cattle
CC   breeding, preparation of constructs for use in cattle gene expression, or
CC   for genetically improving cattle. The present sequence is one of the
CC   15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
CC   present sequence was not shown in the specification but was obtained in
CC   electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137139
XX
SQ   Sequence 473 BP; 129 A; 109 C; 140 G; 94 T; 0 U; 1 Other;

Alignment Scores:

Pred. No.:	1.5e-29	Length:	473
Score:	434.00	Matches:	79
Percent Similarity:	88.78%	Conservative:	8
Best Local Similarity:	80.61%	Mismatches:	11
Query Match:	28.07%	Indels:	0
DB:	7	Gaps:	0

US-10-054-313-1 (1-286) x ABX41815 (1-473)

```
Qy   189  AlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyr 208
          ||:||||||||||||||||||||||
Db   79   GCAGCCTGCAAAGCCATAGAGCAAGCAAGTTAAGGCTCAGGACATCACTAAGCTGGTTCTCTAC 138
```

```
Qy    209 ThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsn 228
          |||||||||||||||||||||||||||||||||||||||||||||||||::::|||
Db    139 ACAGACAGCATGTTTACCATCAATGGCATCACCAACTGGGTGAAGGGCTGGAAGCAGAAT 198

Qy    229 GlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGlu 248
          ||||||::::|||    |||||||||||||||||||||    ||||||
Db    199 GGCTGGAGGACCAGCACCGGGAAGGAGGTGACCAACAAGGAGGACTTCGCGGAGCTGGAG 258

Qy    249 ArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIle 268
          |||    :::|||||||||||||||||||||||||||||||||||||||||||||
Db    259 CGGCTGGGCGCGGGGCATGGACATTCAGTGGACGTTCCTGGCCATTCGGGGTTTAAA 318

Qy    269 GlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAlaLysGlnSerGluAsp 286
          ||||||||||::::|||||||||||||::::||| ||| |||
Db    319 GGCAATGAAAAGGCGGACACGACTATCGAGAGAGGCTCGAAACACTCCGCAGAC 372
```

Search completed: March 22, 2004, 22:42:19
Job time : 337 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 22, 2004, 22:32:11 ; Search time 70 Seconds
                                   (without alignments)
                                   2267.371 Million cell updates/sec

Title: US-10-054-313-1
Perfect score: 1546
Sequence: 1 MSWLLFLAHRVALAALPCR........FIGNEEADRLAREGAKQSED 286

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext   0.5
               Ygapop 10.0 , Ygapext   0.5
               Fgapop 6.0 , Fgapext   7.0
               Delop   6.0 , Delext   7.0

Searched: 682709 seqs, 277747446 residues

Total number of hits satisfying chosen parameters: 136518

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool/US10054313/runat_22032004_140416_28103/app_query.fasta_1.455
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCU=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054313 @CGN_1_1_44 @runat_22032004_140416_28103 -NCPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Issued_Patents_NA:*
         1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
         2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
         3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
         4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
         5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
         6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
          score greater than or equal to the score of the result being printed,
          and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	272	17.6	1326	4	US-09-540-236-1702	Sequence 1702, Ap
2	272	17.6	119211	4	US-09-596-002-40	Sequence 40, Appl
3	268	17.3	231	4	US-09-622-976-14380	Sequence 14380, A
4	241	15.6	1275	4	US-09-252-991A-13638	Sequence 13638, A
5	236.5	15.3	1830121	4	US-09-557-884-1	Sequence 1, Appli
6	236.5	15.3	1830121	4	US-09-643-990A-1	Sequence 1, Appli
7	235.5	15.2	801	4	US-09-199-637A-154	Sequence 154, App
8	235.5	15.2	2048	4	US-09-199-637A-136	Sequence 136, App
9	233.5	15.1	570	4	US-09-252-991A-13802	Sequence 13802, A
10	217	14.0	1509	4	US-09-328-352-1765	Sequence 1765, Ap
11	208.5	13.5	603	4	US-09-489-039A-5519	Sequence 5519, Ap
12	194.5	12.6	489	4	US-09-543-681A-2007	Sequence 2007, Ap

13	11.3	174	501	1	US-08-163-181-3	Sequence 3, Appli
14	11.3	174	501	1	US-08-465-161-3	Sequence 3, Appli
15	11.3	174	552	1	US-08-163-181-4	Sequence 4, Appli
16	11.3	174	552	1	US-08-465-161-4	Sequence 4, Appli
17	10.3	160	8088	4	US-08-716-351A-1	Sequence 1, Appli
18	10.3	160	8535	3	US-08-716-351A-1	Sequence 1, Appli
19	10.2	157	8060	3	US-08-766-528-1	Sequence 1, Appli
20	10.1	156	8202	1	US-08-258-420-13	Sequence 13, Appl
21	10.0	154	2037	4	US-09-397-955C-20	Sequence 20, Appl
22	10.0	154	2037	4	US-09-397-955C-21	Sequence 21, Appl
23	10.0	154	2336	4	US-09-397-955C-18	Sequence 18, Appl
24	10.0	154	3612	3	US-09-265-013-3	Sequence 3, Appli
25	10.0	154	7308	3	US-09-011-745-3	Sequence 3, Appli
26	10.0	154	7308	3	US-09-011-745-4	Sequence 4, Appli
27	10.0	154	7616	3	US-09-011-745-2	Sequence 2, Appli
28	10.0	154	8332	3	US-08-850-961-1	Sequence 1, Appli
29	10.0	154	8332	4	US-09-479-776-1	Sequence 1, Appli
30	10.0	154	8332	4	US-09-309-572-11	Sequence 11, Appl
31	10.0	154	8332	4	US-09-315-127-1	Sequence 1, Appli
32	10.0	154	8332	4	US-09-265-013-1	Sequence 1, Appli
33	10.0	154	8332	4	US-09-554-572-25	Sequence 25, Appl
34	10.0	154	8332	4	US-09-718-096-11	Sequence 11, Appl
35	9.9	153	6363	4	US-08-929-967-6	Sequence 6, Appli
36	9.8	152	8323	2	US-08-110-300A-8	Sequence 8, Appli
37	9.8	152	8323	2	US-08-886-642-8	Sequence 8, Appli
38	9.8	152	8323	4	US-09-433-322B-2	Sequence 2, Appli
39	9.8	152	8323	5	PCT-US93-08041-8	Sequence 8, Appli
40	9.8	152	10367	1	US-08-110-300A-9	Sequence 9, Appli
41	9.8	152	10367	2	US-08-886-642-9	Sequence 9, Appli
42	9.8	152	10367	5	PCT-US93-08041-9	Sequence 9, Appli
43	9.6	149	1953	4	US-09-397-955C-19	Sequence 19, Appl
44	9.6	149	8132	3	US-08-766-528-3	Sequence 3, Appli
45	9.3	144	8655	3	US-09-075-272-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-09-540-236-1702
; Sequence 1702, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CAT
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 1702
; LENGTH: 1326
; TYPE: DNA
; ORGANISM: M.catarrhalis
US-09-540-236-1702

Alignment Scores:
Pred. No.:   1.26e-20   Length: 1326
Score: 272.00   Matches: 62
Percent Similarity: 52.23%   Conservative: 20
Best Local Similarity: 39.49%   Mismatches: 51
Query Match: 17.59%   Indels: 24
DB: 4   Gaps: 5

US-10-054-313-1 (1-286) x US-09-540-236-1702 (1-1326)

Qy   136 MetGlyAspPheValValValTyrThrAspGlyCysCysSerAsnGlyArgArgLys 155
                ||| | ||||||||||| ||| ||||:::::
Db   28 ATGAGCCAACTTAATATCGCTTATACCGACGGCGCCTGTAAAGGCAATGGAAAACAAGGC 87

Qy   156 ProArgAla---GlyIleGlyValTyr------------------------ 163
                |||
Db   88 GTATCTGCAGGCGGTTGGGGCGTGTATTTGCATTATTTCAATGGCGATGAGCGGCATCTG 147

```
RESULT 2
US-09-596-002-40/c
; Sequence 40, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
;   APPLICANT: Lagace, Robert, E.
;   APPLICANT: Paterson, Chandra
;   APPLICANT: Berg, Kim, L.
;   TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
;   FILE REFERENCE: PM-0008-4 US
;   CURRENT APPLICATION NUMBER: US/09/596,002
;   CURRENT FILING DATE: 2000-06-16
;   PRIOR APPLICATION NUMBER: 60/140,121
;   PRIOR FILING DATE: 1999-06-18
;   NUMBER OF SEQ ID NOS: 41
;   SOFTWARE: PERL Program
; SEQ ID NO 40
;   LENGTH: 119211
;   TYPE: DNA
;   ORGANISM: M. catarrhalis
;   FEATURE:
;     NAME/KEY: misc_feature
;     OTHER INFORMATION: Incyte template ID No. 6632636 40
;   PUBLICATION INFORMATION:
US-09-596-002-40

Alignment Scores:
Pred. No.:          1.24e-17          Length:        119211
Score:              272.00            Matches:       62
Percent Similarity: 52.23%            Conservative:  20
Best Local Similarity: 39.49%         Mismatches:    51
Query Match:        17.59%            Indels:        24
DB: 4                                 Gaps:          5

US-10-054-313-1 (1-286) x US-09-596-002-40 (1-119211)

Qy   136 MetGlyAspPheValValTyrThrAspSerGlyCysSerSerAsnGlyArgArgLys 155
             :::
Db 45182 ATGAGCCAACTAATATCGTTATACCGACGGCGCCTGTAAAGGCCATGGAAAACAAGGC 45123

Qy   156 ProArgAla---GlyIleGlyValTyr---------- 163
         |||||  |||
Db 45122 GTATCTGCAGGCGGTTGGGCGTGTATTTGCATTATTTCAATGGCGATGAGCGGCATCTG 45063

Qy   164 TrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGln 183
         |||||
Db 45062 TGGGGC---GGTGAGCCT----------- 45033

Qy   184 ArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsn 203
         |||:||||
Db 45032 CGCATGGAATTGATGGCTGCCATCACAGCTTTGGAGGCAACGGCCTGCACAG-----ATT 44979

Qy   204 LysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGln 223
         ||||||||||
Db 44978 CCTCTGCAACTTGGACAGATTCAGGTATGTTAAGATGGCATAACTCAGTGGATTGGC 44919

Qy   224 GlyTrpLysAsnGlyIleTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAsp 243
         ||||||||
Db 44918 GGTTGGAAGTTGCGTGGTGGAAAAAAGCAGATGGCAAGCCTGTCTTAAATCAAGACCTA 44859

Qy   244 PheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGly 263
         |||:
Db 44858 TGGCAACGACGATTGGATCAGCTGACCCAAAAATCCATCATTGATTGGCAATGGATCAAAGGC 44799

Qy   264 HisSerGlyPheIleGlyAsnGluGluAsnAspArgLeuAlaArgGluGly 280
         ||||||:::||||
Db 44798 CATGCTGGTCATGCTGCAATGAGATGGCAGACCAACTGCCAATAAAGGC 44748


RESULT 3
US-09-621-976-14380
; Sequence 14380, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
;   APPLICANT: Dumas Milne Edwards, J.B.
;   APPLICANT: Jobert, S.
;   APPLICANT: Giordano, J.Y.
;   TITLE OF INVENTION: ESTs and Encoded Human Proteins.
;   FILE REFERENCE: GENSET.054PR2
;   CURRENT APPLICATION NUMBER: US/09/621,976
;   CURRENT FILING DATE: 2000-07-21
;   NUMBER OF SEQ ID NOS: 19335
;   SOFTWARE: Patent.pm
; SEQ ID NO 14380
;   LENGTH: 231
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-621-976-14380

Alignment Scores:
Pred. No.:             2.41e-21       Length:        231
Score:                 268.00         Matches:       63
Percent Similarity:    94.03%         Conservative:  0
Best Local Similarity: 94.03%         Mismatches:    4
Query Match:           17.34%         Indels:        3
DB:                                   Gaps:          0

US-10-054-313-1 (1-286) x US-09-621-976-14380 (1-231)

Qy     1 MetSerTrpLeuLeuPheLeuValHisArgValAlaLeuAlaLeuProCysArgArg 20
Db    33 ATGAGCTGGCTTCTGTTCCTGGCCCCACAGAGTCGGCCGCCTTGGCCGTCGCCGCGGC 92

Qy    21 GlySerArgGlyMetPheTyrAlaValArgGlyLysThrGlyLysValPhe 40
Db    93 GGGCTCTGGCGCGG--TTCGGGGATGTTCTATGCCGTGAG-AGGGGCCCAAGACCGGGGTCTTT 150

Qy    41 LeuThrTrpAsnGluCysArgAlaGlnGlyAlaAspArgPheProAlaAlaArgPheLysLys 60
Db   151 CTGACCTGGAATGAGTGCAGAGCGCAA-GTGGACCGGTTCCTACTGCCCAGATTTAAGAAG 209

Qy    61 PheAlaThrGluAspGluAla 67
Db   210 TTTGCCACAGAGAGATGAGCGCC 230


RESULT 4
US-09-252-991A-13638
; Sequence 13638, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;   APPLICANT: Rubenfield et al.
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;   TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
```

```
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13638
; LENGTH: 1275
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13638

Alignment Scores:
Pred. No.:                 3.38e-17      Length:            1275
Score:                     241.00        Matches:           92
Percent Similarity:        35.96%        Conservative:      36
Best Local Similarity:     25.84%        Mismatches:        126
Query Match:               15.59%        Indels:            102
DB:                                      Gaps:              14

US-10-054-313-1 (1-286) x US-09-252-991A-13638 (1-1275)
```

```
Qy     9  HisArgValAlaLeuAlaAlaLeuProCysArgArgGlySerArgArgGlyPheGlyMetPhe    28
             ||||||    ||   ||    |   ||   |||::     |      ::
Db   259  CACCGGGCAGATTCAGCGCGG-----TGTGCGCCTGGGGCGCCGCCGCTGCCGGGGCGTGGA   312

Qy    29  TyrAlaValArgArgGlyArgLysThrGlyValPheLeuThrTrpAsnGluCysArgAla     48
             :::  :||||||||||    |        |    |  ||
Db   313  CATCGCCTGCGAAGAGGGCGC--------CTGGCGCCTCAGCGAACAATGCCGGCGGA   360

Qy    49  GlnValAspArgPheProAlaAlaArg-----------PheLysLysPheAlaThrGluAsp   65
                              ||||              ||  |||
Db   361  -------CGTGGTCCTGCGCAACACGGCCTGGATTTCTGCCTGTCGCCTCACCGTCT   411

Qy    66  GluAlaTrpAla---PheValArgLysSerAlaSerProGluValSerGluGlyHisGlu     84
             :::        |    |||                   ::  |      |  |
Db   412  CCTGCGCGAAGCCGCGCGTACCGTTCGTCCGGGGCGGCCACCTGCTGTGATCGGCATCAA   471

Qy    85  AsnGlnHis---------------------------------------GlyGln     89
             |||                                          ||||
Db   472  CCCATGAGCGCTGTGGGGCATCCGTCATTATTTCGCCCGGGATGCCTTGCGCCAGGCCCG   531

Qy    90  GluSerGluAlaLysProGlyGlyLysArgLeuArgGluProLeuAspGlyHisGlu   109
             |||         ||||||||||   |  |    |     ||||  :  |||
Db   532  CTGCATTCCTCCGTCGCGGCGGCCTGCGACTGGCTCAACCTGCTGGCGTTCGCGGCTGAGAA   591

Qy   110  SerAla------------------------------------------   111
             :::||
Db   592  ACGGCGCTTCGGGCTGCTATCGTCCGCCGCTTGCGTCGGCAGCCTGGCAATCGGCGCCTGGC   651

Qy   112  GlnProTyrAlaLysHisMetLys----------------------------   119
             :::                   :::
Db   652  TCGCCTGCGAGCGCGCTGGGGCGCACCGCCTGGCAGTCTTCGGGGCGCCTTCTATCTATTGGT   711

Qy   120  ----------------------ProSerValGluProAlaProValSerArgAspThr-------   132
                                   ||          |
Db   712  GGCACGCAAGCTGCTGTCGTCGGGGTTGCGCCCGTTGCGCAGAGCAAGCCGCAACCGGCGGCG   771

Qy   133  ------------------------PheSerTyrMetGly---   137
                                    |||       :::|||
Db   772  TCAGCTGGTGCCCATGCCGGTGGGCGAAAGTCAGCCGGCGGAGATTTCCGAAATTTAGGCATG   831

Qy   138  ----------AspPheValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArg   154
                      :::  |||||:||||||::||:||      |    ||
Db   832  ACAGATAAAGAAGAACAGGTAGTGATCTATACCGACGGGTGCTGCAGTGGCAACCCTGGGCGC   891

Qy   155  LysProAlaArgAlaGlyIleGlyValTyr----------------TrpGlyProGly   167
             :::|||                 ||:   |||
Db   892  GGGCGGCTGGGGGGCCGGTTGCTCCTCTACAAGGGCCACGGCGAGCGTTTGGGGC--GGC   948

Qy   168  HisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrArgAsnGlnArgAlaGluIle   187

Db   949  GAGCCG                                                  -GACACCACCAACACCGGCATGGAACTG   981
             |||                                                     ||||||||     |||::

Db   188  HisAlaAlaCysLysLysAlaIleGluGlnIleAlaLysThrGlnAsnIleAsnLysLeuValLeu   207
             ||||||::      :::|                :  |::|||
Qy   982  ATGGCGGGGATCAGGCGGCTGGCGCACTCAGCGGTCCTGTCCGATCCGTCTGATC--   1038

Qy   208  TyrTyrThrAspSerMetPheThrIleAlaAsnGlyIleThrAsnTrpValGlnGlyTrpLysTrpLys   227
             |||||||                |||||||||
Db  1039  ---ACCGACTCGGAAATACGTGATCGATGCGCCGGCATCACCGAATGGTTGCGAACTGGAAGAAG   1095

Qy   228  AsnGlyTyrTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu   247
             ||||||||||||||||||:: :::        |||    :::  |||
Db  1096  CGCGGCTGGAAGACCGCCAAGCAAGCAGCCGGTCAAGAATGCCGACCTCTGGCAGCGCCCTG   1155

Qy   248  GluArgLeuThrGlnGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe   267
             :::                              :::::|||
Db  1156  GATGAGCAGGTCGCCCGGCACCAGGTGGAGTGGCAGTGGGTCCGCGGGCATACCGGCGAT   1215

Qy   268  IleGlyAsnGluGlyAlaAspSerArgLeuAlaArgGluGluGluGlyLysAlaLysGln   283
             |||                              ||||  :::|||
Db  1216  CCCGGGCAACGACGCGGGCGGCCCAACCGTTGGCCAACCAGTTGGCGCGTCGCCGAA   1263

RESULT 5
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
;   APPLICANT: Fleischmann et al.
;   TITLE OF INVENTION: The Nucleotide sequence of
                         the Haemophilus influenzae Rd Genome, Fragments
                         Thereof, and Uses Thereof
;   NUMBER OF SEQUENCES: 1
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Human Genome Sciences, Inc.
;     STREET: 9410 Key West Avenue
;     CITY: Rockville
;     STATE: MD
;     COUNTRY: USA
;     ZIP: 20850
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: 3 1/2 inch diskette
;     COMPUTER: Dell Pentium
;     OPERATING SYSTEM: MS DOS v6.22
;     SOFTWARE: ASCII Text
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/557,884
;     FILING DATE: 25-Apr-2000
;     CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/476,102
;     FILING DATE: JUN-5-1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Michelle S. Marks
;     REGISTRATION NUMBER: 41,971
;     REFERENCE/DOCKET NUMBER: PB186P3
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 301-309-8504
;     TELEFAX: 301-309-8439
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1830121 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: double
;     TOPOLOGY: linear
;   SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Alignment Scores:
Pred. No.:                 7.4e-12      Length:            1830121
Score:                     236.50       Matches:           54
Percent Similarity:        49.68%       Conservative:      23
Best Local Similarity:     34.84%       Mismatches:        49
```

NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Alignment Scores:
Pred. No.:   7.4e-12   Length: 1830121
Score: 236.50   Matches: 54
Percent Similarity: 49.68%   Conservative: 23
Best Local Similarity: 34.84%   Mismatches: 49
Query Match: 15.30%   Indels: 29
DB: 4   Gaps: 4

US-10-054-313-1 (1-286) x US-09-643-990A-1 (1-1830121)

Qy   140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
                :::   ::::::|||||||| ::: ||||||
Db 152742 ATTGAAATTTTACTGATGGATCTTGCTTAGGTAAT----------CCAGGGGCGGGC 152789

Qy   160 IleGlyValTyrTrpGlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
                                                       :::|||||| ||| :::
Db 152790 ------------------------------------GGAATTGGTGCCGTATTGCGTTATAAA 152816

Qy   180 Gln--------------------------------ThrAsnGlnArgAlaGluIle 187
                ||| |||||||||| ||||||
Db 152817 CAACATGAAAAAACACTCTCCAAAGGCTATTTCCAAACCACCAATAATCGAATGGAATTA 152876

Qy   188 HisAlaAlaCysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu 207
                |||||| |||||| :::   :::----ATCACGCTT
Db 152877 CGCGCTGTCATTGAAGCATTAAATACATTAAAAGAACCTGCTTG-----ATCACGCTT 152930

Qy   208 TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys 227
                :::||:|| |||||||| ::::::|| ||||||
Db 152931 TATAGTGATGATAGCCAATATATGAAAAATGCATAACCAAATGGATCTTTAACTGAAAAA 152990

Qy   228 AsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu 247
                ||||||| ||||||:::||||||| :::::||||||||||
Db 152991 AATAATTGGAAAGCAAGTTCTGGAAAGCCTGTAAAAAACCAAGATTTATGGATAGCCTTA 153050

Qy   248 GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe 267
                ::: |||||||| ||:::||
Db 153051 GATGAATCCATCCAACGTCATAAAATTAATTGGCAATGGTAAAAGGCCATGCTGGACAC 153110

Qy   268 IleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAlaLys 282
                ||||||||||||||||:::||:::||||||
Db 153111 AGAGAAAATGAAATTTGCCGATGAATTAGCAAAAAAAGGGGCAGAA 153155

RESULT 7
US-09-199-637A-154
; Sequence 154, Application US/09199637A
; Patent No. 6355411
GENERAL INFORMATION:
APPLICANT: Ausubel, Frederick
APPLICANT: Goodman, Howard M.
APPLICANT: Rahme, Laurence G.
APPLICANT: Mahajan-Miklos, Shalina
APPLICANT: Tan, Man-Wah
APPLICANT: Cao, Hui
APPLICANT: Dreikard, Eliana
APPLICANT: Tsongalis, John
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
TITLE OF INVENTION: SEQUENCES AND USES THEREOF
FILE REFERENCE: 00786/361002

```
;  CURRENT APPLICATION NUMBER: US/09/199,637A
;  CURRENT FILING DATE:  1998-11-25
;  PRIOR APPLICATION NUMBER:  60/066,517
;  PRIOR FILING DATE: 1997-11-25
;  NUMBER OF SEQ ID NOS:  437
;  SOFTWARE:  FastSEQ for Windows Version 4.0
;  SEQ ID NO 154
;    LENGTH: 801
;    TYPE: DNA
;    ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-154

Alignment Scores:
Pred. No.:                6.79e-17        Length:            801
Score:                    235.50          Matches:            60
Percent Similarity:       49.12%          Conservative:       24
Best Local Similarity:    35.09%          Mismatches:         64
Query Match:              15.23%          Indels:             24
DB:                       4               Gaps:                5

US-10-054-313-1 (1-286) x US-09-199-637A-154 (1-801)

Qy    124  ProAlaProProValSerArgAspThrPheSerTyrMetGly-----------AspPhe  139
                 ||||||               |||||:|          :::
Db    314  CCGGTGGCGAAAGTCAGCCGGCG-AGATTCCGAAATTTAGGCATGACAGATAAAGAACAG  372

Qy    140  ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly  159
                 ||||:::::||| ||||||||  ||||||||||||||||||||
Db    373  GTAGTGATCTATACCGACGGGCGCCTGCAAGGGCAACCCTGGGCGCGGCGCTGGGGGGCG  432

Qy    160  IleGlyValTyr-------------------TrpGlyProGlyHisProLeuAsnVal  172
                 ::: :::||||                     ||||||           |||||||
Db    433  TTGCTCCTCTACAAGGGCGCCGAGCGAGCTTGGGGC---GGCGAGCCG----------  480

Qy    173  GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLys  192
                 ||||||||||||||||    ||||||||||||||||||       :||
Db    481  -----------------------GACACCACCAACACCGCATGGCGGCGGCGGCGATCCAG  522

Qy    193  AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet  212
                 ||||:  |               :::|||||||:          :::
Db    523  GGCGCTGGCGCCACTCAAGCGTTCCTGTCCGATCCGTCTGATC-----ACCGACTCGGAA  576

Qy    213  PheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThr  232
                 :::   |||||||  |||||          |||||||||||||||||
Db    577  TACGTGATGCGGCGCATCACCGAATGGTTGCCGAACTGGTTGCCGAACTGGCTGGAAGACC  636

Qy    233  SerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGln  252
                 ::::: :::||       |||    ||    ||||:::
Db    637  GCCAGCAAGCAGCCTGTCAAGAATGCCGACCTCTGGCAGGCCCGATGAACAGGTCGCC  696

Qy    253  GlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGlu  272
                 ||||||||||||||| ||:::||||||||:::||
Db    697  CGGCACCAGTGGAGTGGCAGTGGCATACCGGCGCACCCCGGCCAACGAGCGG  756

Qy    273  AlaAspArgLeuAlaArgGluGlyAlaLysGln  283
                 ||||||||||:::|||       :::
Db    757  GCCGACCAGTTGGCCAACCGTGGGCGTCGCCGAA  789
```

RESULT 8
US-09-199-637A-136/c
; Sequence 136, Application US/09199637A
; Patent No. 6355411
; GENERAL INFORMATION:
; APPLICANT: Ausubel, Frederick
; APPLICANT: Goodman, Howard M.
; APPLICANT: Rahme, Laurence G.
; APPLICANT: Mahajan-Miklos, Shalina
; APPLICANT: Tan, Man-Wah
; APPLICANT: Cao, Hui
; APPLICANT: Drenkard, Eliana
; APPLICANT: Tsongalis, John
; TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
; TITLE OF INVENTION: SEQUENCES AND USES THEREOF

```
;  FILE REFERENCE:  00786/361002
;  CURRENT APPLICATION NUMBER: US/09/199,637A
;  CURRENT FILING DATE:  1998-11-25
;  PRIOR APPLICATION NUMBER:  60/066,517
;  PRIOR FILING DATE: 1997-11-25
;  NUMBER OF SEQ ID NOS:  437
;  SOFTWARE:  FastSEQ for Windows Version 4.0
;  SEQ ID NO 136
;    LENGTH: 2048
;    TYPE: DNA
;    ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-136

Alignment Scores:
Pred. No.:                2.86e-16        Length:           2048
Score:                    235.50          Matches:            60
Percent Similarity:       49.12%          Conservative:       24
Best Local Similarity:    35.09%          Mismatches:         64
Query Match:              15.23%          Indels:             24
DB:                       4               Gaps:                5

US-10-054-313-1 (1-286) x US-09-199-637A-136  (1-2048)

Qy    124  ProAlaProProValSerArgAspThrPheSerTyrMetGly-----------AspPhe  139
                 |||               |||||:|          :::
Db    648  CCGGTGGCGAAAGTCAGCCGGCG-AGATTCCGAAATTTAGGCATGACAGATAAAGAACAG  590

Qy    140  ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly  159
                 ||||:::::||| ||||||||  ||||||||||||||||||||
Db    589  GTAGTGATCTATACCGACGGGCGCCTGCAAGGGCAACCCTGGGCGCGGCGCTGGGGGGCG  530

Qy    160  IleGlyValTyr-------------------TrpGlyProGlyHisProLeuAsnVal  172
                 ::: :::||||                     ||||||           |||||||
Db    529  TTGCTCCTCTACAAGGGCGCCGAGCGAGCTTGGGGC---GGCGAGCCG----------  482

Qy    173  GlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLys  192
                 ||||||||||||||||    ||||||||||||||||||       :||
Db    481  -----------------------GACACCACCAACACCGCATGGCGGCGGCGGCGATCCAG  440

Qy    193  AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet  212
                 ||||:  |               :::|||||||:          :::
Db    439  GGCGCTGGCGCCACTCAAGCGTTCCTGTCCGATCCGTCTGATC-----ACCGACTCGGAA  386

Qy    213  PheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLysAsnGlyTrpLysThr  232
                 :::   |||||||  |||||          |||||||||||||||||
Db    385  TACGTGATGCGGCGCATCACCGAATGGTTGCCGAACTGGTTGCCGAACTGGCTGGAAGACC  326

Qy    233  SerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGln  252
                 ::::: :::||       |||    ||    ||||:::
Db    325  GCCAGCAAGCAGCCTGTCAAGAATGCCGACCTCTGGCAGGCCCGATGAACAGGTCGCC  266

Qy    253  GlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGlu  272
                 ||||||||||||||| ||:::||||||||:::||
Db    265  CGGCACCAGTGGAGTGGCAGTGGCATACCGGCGCACCCCGGCCAACGAGCGG  206

Qy    273  AlaAspArgLeuAlaArgGluGlyAlaLysGln  283
                 ||||||||||:::|||       :::
Db    205  GCCGACCAGTTGGCCAACCGTGGGCGTCGCCGAA  173
```

RESULT 9
US-09-252-991A-13802/c
; Sequence 13802, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190

```
;   PRIOR FILING DATE: 1998-07-27
;   NUMBER OF SEQ ID NOS: 33142
;   SEQ ID NO 13802
;   LENGTH: 570
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13802

Alignment Scores:
Pred. No.:                6.74e-17      Length:          570
Score:                    233.50        Matches:         54
Percent Similarity:       50.33%        Conservative:    22
Best Local Similarity:    35.76%        Mismatches:      56
Query Match:              15.10%        Indels:          19
DB:                                     Gaps:            4

US-10-054-313-1 (1-286)  x US-09-252-991A-13802 (1-570)

Qy    140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
          |||||                                    ::: :::        ||||
Db    546 GTAGTGATCTATACCGACGGCGCCTGCAAGGGCCAACCCTGGCGCGCGGCTGGGGGGCG 487

Qy    160 IleGlyValTyr------------------TrpGlyProGlyHisProLeuAsnVal 172
          ||: ::|||                            ||||:||| ||
Db    486 TTGCTCCCTCACAAGGGCGCGAGCGAGAGCTTTGGGGC---GGCGAGCCG---------- 439

Qy    173 GlyIleArgLeuProGlyArgGlnThrArgGlnAlaArgAlaGlyIleHisAlaAlaCysLys 192
          |||||                 ::: :::           ||||||           ::: |||
Db    438 -----------------GACACCACCAACAACCGCATGAACTGATGGCGGCGATCCAG 397

Qy    193 AlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMet 212
          ||||:                    ::|||:||||
Db    396 GCGCTGGCGGCCACTCAAGCGTTCCGTCCGATCCGTTCGATC------ACCGACTCGGAA 343

Qy    213 PheThrIleAsnGlyIleThrAsnTrpValGlnGlyThrProLysLysAsnGlyIleThrProLysLysThr 232
          ||||| ::: ||:                    |||||||||||||||
Db    342 TACGTGATGCGCGGCCATCACCGAATGGTTGCCGAACTGGAAGAAGCGCGGCTGGAAGACC 283

Qy    233 SerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeuGluLeuThrGln 252
          ||||||                ::: :::
Db    282 GCCAGCAAGCAGCAGCCGGTCAAGAATGCCGACCTCTGGCAGCCCTCGATGAGCAGGTCGCC 223

Qy    253 GlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGlu 272
          ||||||                 ::: :::|||
Db    222 CGGCACCAGGTGGAGTGGCAGTGGGTTCCGCCGGGCATACCCGGCAACCGAGCGG 163

Qy    273 AlaAspArgLeuAlaArgGluGluGlyAlaLysGln 283
          ||||||                   :::
Db    162 GCCGACCAGTTGGCCAACCGTGGGCGTCGCCGAA 130
```

```
RESULT 10
US-09-328-352-1765
;   Sequence 1765, Application US/09328352
;   Patent No. 6562958
;   GENERAL INFORMATION:
;   APPLICANT: Gary L. Breton et al.
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;   TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;   FILE REFERENCE: GTC99-03PA
;   CURRENT APPLICATION NUMBER: US/09/328,352
;   CURRENT FILING DATE: 1999-06-04
;   NUMBER OF SEQ ID NOS: 8252
;   SEQ ID NO 1765
;   LENGTH: 1509
;   TYPE: DNA
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-1765

Alignment Scores:
Pred. No.:                2.07e-14      Length:          1509
Score:                    217.00        Matches:         52
Percent Similarity:       46.00%        Conservative:    24
Best Local Similarity:    31.52%        Mismatches:      53
```

```
Query Match:              14.04%        Indels:          36
DB:                                     Gaps:            6

US-10-054-313-1 (1-286)  x US-09-328-352-1765 (1-1509)

Qy    133 PheSerTyrMetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGly 152
          |||  :::|||                  ::: :::                      |||
Db    121 TTTCGATTATGTCTCAAACAATCACCACTGTATGTTGATGGTGCATGTCGTGCGCAAT-- 177

Qy    153 ArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyPro---------------- 166
          |||                             ||| :::
Db    178 ----------------CCGGGTTTAGGC--GGTTGGGGTGCATATGTCATTACCGAGCAG 219

Qy    167 -------------------GlyHisProLeuAsnValGlyIleArgLeuProGlyArg 179
                             ||||:||| |||
Db    220 GGTGAACATAAATTATTGGCGGTGAACCC----------------------GAC 252

Qy    180 GlnThrArgGlnAlaArgAlaGlyIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThr 199
          ||||||                 ::: :::           :::
Db    253 ACAACAAATAACCGTATGGCAGCTTACTGCGGTCATTGAAGGAGAATTTCTTTTGCCCTCCT 312

Qy    200 GlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThr 219
          ||||||                              :::
Db    313 GAT-------GCTCAATTAATTGTCTGGACAGACTCAAATTATGTGAAGCAAGGTATTACA 366

Qy    220 AsnTrpValGlnGlyThrProLysLysAsnGlyIleThrSerAlaGlyLysGluValIle 239
          |||||                 :::
Db    367 GAATGGATTCATGGTTGGAAAAAAAGAAAAACTGG-----------AAAGATGTTAAA 411

Qy    240 AsnLysGluAspPheValAlaLeuGluLeuThrGlnGlyMetAspIleGlnTrpMet 259
          |||  :::                 :::
Db    412 AATCCTGACCTATGGCAAAAACTCGATGCCGTTTGTGGGGTAGAAAATTGAATGGAAC 471

Qy    260 HisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGlu 279
          |||||                   |||||
Db    472 TGGATTAAAGGCCATGCAGGACGACATGCAGGTAATGAAAATGGCACCAGCTTGCGAATTTA 531

Qy    280 GlyAlaLysGlnSer 284
          |||||    :::
Db    532 GGTGCAGACAAAACT 546
```

```
RESULT 11
US-09-489-039A-5519
;   Sequence 5519, Application US/09489039A
;   Patent No. 6610836
;   GENERAL INFORMATION:
;   APPLICANT: Gary Breton et. al
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;   TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;   FILE REFERENCE: 2709.2004001
;   CURRENT APPLICATION NUMBER: US/09/489,039A
;   CURRENT FILING DATE: 2000-01-27
;   PRIOR APPLICATION NUMBER: US 60/117,747
;   PRIOR FILING DATE: 1999-01-29
;   NUMBER OF SEQ ID NOS: 14342
;   SEQ ID NO 5519
;   LENGTH: 603
;   TYPE: DNA
;   ORGANISM: Klebsiella pneumoniae
US-09-489-039A-5519

Alignment Scores:
Pred. No.:                4.49e-14      Length:          603
Score:                    208.50        Matches:         54
Percent Similarity:       44.81%        Conservative:    15
Best Local Similarity:    35.06%        Mismatches:      56
Query Match:              13.49%        Indels:          29
DB:                                     Gaps:            4

US-10-054-313-1 (1-286)  x US-09-489-039A-5519 (1-603)

Qy    140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
          |||                              :::::|||||||||||            ||||
```

```
Db  148  GTAGAAATTTCACCGACGGCTCTTGCCTCGGTAATCCA-----                   186
         ||||||
Qy  160  IleGlyValTyrTyrProGly------------------                       167

Db  187  --GGACCTGGGGCCACGGCGCCATCATGCGCTACCGCCAGCACGAA                231
                |||||
Qy  168  HisProLeuAsnValGlyIleArgLeuProGlyArgGlnThrAsnGlnArgAlaGluIle  187

Db  232  AAAACCTTCAGCGCTGGCTACCGTCTG--------ACCACCAATAACCGCATGATGGAACTG  282
              |||::||
Qy  188  HisAlaAlaCysLysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeu  207

Db  283  ATGGCGGCAATTGTGCCCTGGAGCGGTTAAAAGAGACAT-----TGCGAAGTGGTGCTC   336
         ||||||                              :::||:||
Qy  208  TyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTrpLysLys  227

Db  337  AGCACTGACAGCCAGCCAGTATGTTCGTCAGGGGATCACCCAGTGGAAAACTGGAAAAAG  396
         :::
Qy  228  AsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAlaLeu  247

Db  397  CGCGGCTCGAAAACGGCAGAGAAAAAGCCGGTGAAAAATGTCGACCTCTGGCAGCCGCTG  456
         ||||||              :::         :::
Qy  248  GluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGlyPhe  267

Db  457  GACGCCCGCGCCTTGGCCAGCATAAAATTAAATGGGAATGGTAAAGGGCCATGCCGGCCAT  516
         :::                                     ||::||
Qy  268  IleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAla                    281

Db  517  CCTGAAAATGAACGCTGCCGACGAGCTGGCGCGCGGCGGCC                     558
```

RESULT 12
US-09-543-681A-2007
; Sequence 2007, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABIL
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 2007
; LENGTH: 489
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-2007

Alignment Scores:
Pred. No.:
Score: 194.50
Percent Similarity: 45.34%
Best Local Similarity: 31.06%
Query Match: 12.58%
DB:

Length: 489
Matches: 50
Conservative: 23
Mismatches: 59
Indels: 29
Gaps: 4

Pred. No.:   1.18e-12

US-10-054-313-1 (1-286) x US-09-543-681A-2007 (1-489)

```
Qy  134  SerTyrMetGlyAspPheAlaValValTyrThrAspGlyCysSerSerAsnGlyArg    153
         :::::||                                 :::::||
Db    4  GCCTTATGCACAAGCAGCAGTAGAAATATTCACCGATGGTTCATGTTTAGGCACCCA---  60

Qy  154  ArgLysProArgAlaGlyAlaGlyIleGlyValTyrTrpGlyHisProLeuAsnValGly  173
                                                  |||||||
Db   61  -----------------------GGTCCTGGT--------GGTTATGGT            78

Qy  174  IleArgLeuProGlyArgGln------------------------Thr            181
         ||| :::
Db   79  GCAATTTTACGCTACCAACAGCATGAGAAAAACCCTTAGTGAAGGTTTTTTATGACCACC  138

Qy  182  AsnGlnArgAlaGluIleHisAlaAlaCysLysLysAlaIleAlaGlnAlaLysThrGlnAsn  201
```

```
Db  139  AATAACCGCATGGAACTGCCCTTGCTGCTCATCGTATCGTCCTATCGTTAGAGGCGTTAAAATTCCCT-  195
         |||                           ||::||               |||
Qy  202  IleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrp  221

Db  196  ---TGTAAAATTACACTGACTACGGATAGCCAGTATGTCAGACAGGGAATTACCAAGTGG  252
                                              :::                   |||
Qy  222  ValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLys  241

Db  253  ATACATAGTTGGAAAACACGCCAATGCGTAAAGCAGATAAAAGCCCTGTCGTGCTGAATGTT  312
         :::                                 :::
Qy  242  GluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisVal  261

Db  313  GATTTATGGAAGCGTCTGATAAAGCCATGAAATTGAATGGCGTCATTGGGTT          372
         :::          :::                         ||||||:||
Qy  262  ProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGlyAla  281

Db  373  AAAGGTCATGCAGGCCATGACGAGAAATAGAACGTTGTGATGAACTGGCCAAGCGGCCGCG  432
         ||||||:|
Qy  282  Lys                                                          282
         :::
Db  433  CAA                                                          435
```

RESULT 13
US-08-163-181-3
; Sequence 3, Application US/08163181
; Patent No. 5459055
; GENERAL INFORMATION:
; APPLICANT: Jendrisak, Jerome J.
; APPLICANT: Smith, Robert E.
; APPLICANT: Dahl, Gary A.
; TITLE OF INVENTION: THERMOSTABLE RIBONUCLEASE H
; TITLE OF INVENTION: AND GENETIC CONSTRUCT
; TITLE OF INVENTION: THEREFORE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: U.S.A.
; ZIP: 53202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/163,181
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/815,095
; FILING DATE: 27-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Baker, Jean C.
; REGISTRATION NUMBER: P-35,433
; REFERENCE/DOCKET NUMBER: 31-307-9001-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5709
; TELEFAX: (414) 277-5774
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 501 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-163-181-3   Length: 501

Alignment Scores:
Pred. No.:   2.36e-10

```
Score:                  174.00      Matches:        51
Percent Similarity:     51.35%      Conservative:   25
Best Local Similarity:  34.46%      Mismatches:     59
Query Match:            11.25%      Indels:         14
DB:                                 Gaps:            5

US-10-054-313-1 (1-286) x US-08-163-181-3 (1-501)

Qy  140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
            :::||||||||||      |||   ::: ||    |||
Db   25 GTGGCCCTCTCACCGACGGACGGGGGCCTGCCTGGGAAACCC-CGGGCCCGG-------GGG 74

Qy  160 IleGlyValTyrTrpGlyPro--------GlyHisProLeuAsnValGlyIleArgLeu 176
            ::||                      ||| :::|||
Db   75 GTGGGC--------GGCCCTCCTCCGCTTCCACGCCCACGAGAAGCTCCTCTCCGGGGG 125

Qy  177 ProGlyArgGln-ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGl 196
            |||||       ||||   :::   |||    ::: :::::
Db  126 AGAGGCCTGCACCACCAACAACCGCCATGGAGCTCAAGGCGCCATAGAGGGCCTAAAGGC 185

Qy  196 nAlaLysThrGlnAsnIleAsnLysValLeuTyrThrAspSerMetPheThrIleAs 216
            :::::           |||||||       :::
Db  186 CCTCAAGGAGCCT------TGCGAGGTGGACCTCTACACCGACGCCACTACCTCAAGAA 239

Qy  216 nGlyIleThr--AsnTrpValGlnGlyTrpLysLysAsnGlyThrTrpLysThrSerAlaGl 235
            |||||||      :::||||||   ||||    :::::
Db  240 GGCCTTCACCGAGGGCTGGCTGAAGGCTGGGGGAAAAGGGGCTGGGCGGCGGAGGG 299

Qy  235 YLysGluValIleAsnLysGluAspPheValAlaLeuGluThrGlnGlyMetAs 255
            ||::|::|||       :::
Db  300 CAAGCCCGTGAAAAACCGCGACCTCTGGGAGGGCCCTCCTCCTCGGCCCCCCACCG 359

Qy  255 pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspAr 275
            ::::||      |||||||   :::    ||::      :::::
Db  360 GGTGCGCTTCCACTTCGTGAAGGGGCACACGGGCCACCCGGAGAACGAACGGGTGACCG 419

Qy  275 gLeuAlaArgGluGlyAlaLys 282
            |||||||    :::
Db  420 GGAGGCGGAGGCGGCCAGGCCCAG 441


RESULT 14
US-08-465-161-3
; Sequence 3, Application US/08465161
; Patent No. 5500370
GENERAL INFORMATION:
  APPLICANT: Jendrisak, Jerome J.
  APPLICANT: Smith, Robert E.
  APPLICANT: Dahl, Gary A.
  TITLE OF INVENTION: THERMOSTABLE RIBONUCLEASE H AND GENETIC CONSTRUCTS THEREFORE
  NUMBER OF SEQUENCES: 4
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Quarles & Brady
    STREET: 411 East Wisconsin Avenue
    CITY: Milwaukee
    STATE: Wisconsin
    COUNTRY: U.S.A.
    ZIP: 53202
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version
    SOFTWARE: #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/465,161
    FILING DATE:
    CLASSIFICATION: 435
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US 07/815,095
    FILING DATE: 27-DEC-1991
  ATTORNEY/AGENT INFORMATION:
    NAME: Baker, Jean C.
    REGISTRATION NUMBER: P-35,433
```

```
;   REFERENCE/DOCKET NUMBER:  31-307-9001-1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (414) 277-5709
;     TELEFAX:    (414) 277-5774
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  501 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-465-161-3

Alignment Scores:
Pred. No.:
Score:                  2.36e-10    Length:        501
Percent Similarity:     174.00      Matches:        51
Best Local Similarity:  51.35%      Conservative:   25
Query Match:            34.46%      Mismatches:     59
DB:                     11.25%      Indels:         14
                                    Gaps:            5

US-10-054-313-1 (1-286) x US-08-465-161-3 (1-501)

Qy  140 ValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly 159
            :::||||||||||      |||   ::: ||    |||
Db   25 GTGGCCCTCTCACCGACGGACGGGGGCCTGCCTGGGAAACCC-CGGGCCCGG-------GGG 74

Qy  160 IleGlyValTyrTrpGlyPro--------GlyHisProLeuAsnValGlyIleArgLeu 176
            ::||                      ||| :::|||
Db   75 GTGGGC--------GGCCCTCCTCCGCTTCCACGCCCACGAGAAGCTCCTCTCCGGGGG 125

Qy  177 ProGlyArgGln-ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGl 196
            |||||       ||||   :::   |||    ::: :::::
Db  126 AGAGGCCTGCACCACCAACAACCGCCATGGAGCTCAAGGCGCCATAGAGGGCCTAAAGGC 185

Qy  196 nAlaLysThrGlnAsnIleAsnLysValLeuTyrThrAspSerMetPheThrIleAs 216
            :::::           |||||||       :::
Db  186 CCTCAAGGAGCCT------TGCGAGGTGGACCTCTACACCGACGCCACTACCTCAAGAA 239

Qy  216 nGlyIleThr--AsnTrpValGlnGlyTrpLysLysAsnGlyThrTrpLysThrSerAlaGl 235
            |||||||      :::||||||   ||||    :::::
Db  240 GGCCTTCACCGAGGGCTGGCTGAAGGCTGGGGGAAAAGGGGCTGGGCGGCGGAGGG 299

Qy  235 YLysGluValIleAsnLysGluAspPheValAlaLeuGluThrGlnGlyMetAs 255
            ||::|::|||       :::
Db  300 CAAGCCCGTGAAAAACCGCGACCTCTGGGAGGGCCCTCCTCCTCGGCCCCCCACCG 359

Qy  255 pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspAr 275
            ::::||      |||||||   :::    ||::      :::::
Db  360 GGTGCGCTTCCACTTCGTGAAGGGGCACACGGGCCACCCGGAGAACGAACGGGTGACCG 419

Qy  275 gLeuAlaArgGluGlyAlaLys 282
            |||||||    :::
Db  420 GGAGGCGGAGGCGGCCAGGCCCAG 441


RESULT 15
US-08-163-181-4
; Sequence 4, Application US/08163181
; Patent No. 5459055
GENERAL INFORMATION:
  APPLICANT: Jendrisak, Jerome J.
  APPLICANT: Smith, Robert E.
  APPLICANT: Dahl, Gary A.
  TITLE OF INVENTION: THERMOSTABLE RIBONUCLEASE H
  TITLE OF INVENTION: AND GENETIC CONSTRUCT
  TITLE OF INVENTION: THEREFORE
  NUMBER OF SEQUENCES: 4
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Quarles & Brady
    STREET: 411 East Wisconsin Avenue
    CITY: Milwaukee
    STATE: Wisconsin
    COUNTRY: U.S.A.
```

```
     ZIP:  53202
COMPUTER READABLE FORM:
     MEDIUM TYPE:  Floppy disk
     COMPUTER:  IBM PC compatible
     OPERATING SYSTEM:  PC-DOS/MS-DOS
     SOFTWARE:  PatentIn Release #1.0, Version
     SOFTWARE:  #1.25
CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  US/08/163,181
     FILING DATE:
     CLASSIFICATION:  435
PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  US 07/815,095
     FILING DATE:  27-DEC-1991
ATTORNEY/AGENT INFORMATION:
     NAME:  Baker, Jean C.
     REGISTRATION NUMBER:  P-35,433
     REFERENCE/DOCKET NUMBER:  31-307-9001-1
TELECOMMUNICATION INFORMATION:
     TELEPHONE:  (414) 277-5709
     TELEFAX:  (414) 277-5774
INFORMATION FOR SEQ ID NO:  4:
SEQUENCE CHARACTERISTICS:
     LENGTH:  552 base pairs
     TYPE:  nucleic acid
     STRANDEDNESS:  double
     TOPOLOGY:  linear
     MOLECULE TYPE:  Other Nucleic Acid
US-08-163-181-4
```

```
Alignment Scores:
Pred. No.:                2.74e-10          Length:         552
Score:                    174.00            Matches:        51
Percent Similarity:       51.35%            Conservative:   25
Best Local Similarity:    34.46%            Mismatches:     59
Query Match:              11.25%            Indels:         14
DB:                       1                 Gaps:           5
```

```
US-10-054-313-1 (1-286) x US-08-163-181-4 (1-552)

Qy   140  ValValValTyrThrArgThrGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGly  159
          :::::::||||||||    ::                          :::       |||
Db    76  GTGGCCCTTTCACCGACGGGGCCGCCTGGGAAACCC-CGGGCCCG--------GGG  125

Qy   160  IleGlyValTyrTrpTrpGlyPro---------GlyHisProLeuAsnValGlyIleArgLeu  176
          :::|||||    ||||||                               :::|||
Db   126  GTGGGC---------GGCCCTCCTCCGGTTCCACGCGCCACGAGAAGCTCCTCTCCGGGGG  176

Qy   177  ProGlyArgGln-ThrArgGlnArgAlaGluIleHisAlaAlaCysGluAlaIleGluGl  196
          ||||||    ||||||         ||:::  |||||   :::  :::::::
Db   177  AGAGGCCTGCACCACCAACCGCATGAGCTCAAGGCGGCCATAGAGGGGCTAAAGGC  236

Qy   196  nAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAs  216
          :::::::                                 :::
Db   237  CCTCAAGGAGCCT------TGCGGGTGGGACCCTCACACCGGAGCAGCCACTACCTCAAGAA  290

Qy   216  nGlyIleThr---AsnTrpValGlnGlyTrpValTrpLysLysAsnGlyTrpLysThrSerAlaGl  235
          |||:::||||||:::||||  :::||||    :::|||:::
Db   291  GGCCTTCACCGACGGGGCTGCTGGAAGGCTGGCGGAAAAAGGGGCTGGCGGACGGCGGAGGG  350

Qy   235  yLysGluValIleAsnLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAs  255
          |||||    ||||||   :::  :::||||||
Db   351  CAAGCCCGTGAAAAACCGCGACCCTCTGGGAGGCCCTCCTCCTCGCCATGGCCCCCACCG  410

Qy   255  pIleGlnTrpMetHisValProGlyHisSerGlyPheIleGlyHisGluGluGlnAlaAspAr  275
          |||||||         ||||||            ||||||
Db   411  GGTGCGGTTCCACTTCGTGAAGGGGCACACGGGGCCACCCGGAGAACGGGGTGGACCG  470

Qy   275  gLeuAlaArgGluGlyAlaLys  282
          |||||||    ||:::
Db   471  GGAGGCGGAGGCGGCCAGGCCCAG  492
```

```
                    Copyright (c) 1993 - 2004 Compugen Ltd.
                                GenCore version 5.1.6

OM protein - nucleic search, using frame_plus_p2n model

Run on:        March 22, 2004, 22:36:31 ;  Search time 292 Seconds
                                           (without alignments)
                                           3623.143 Million cell updates/sec

Title:         US-10-054-313-1
Perfect score: 1546
Sequence:      1 MSWLLFLAHRVALALALPCRR.........FIGNEEADRLAREGAKQSED 286

Scoring table: BLOSUM62
               Xgapop 10.0 ,  Xgapext  0.5
               Ygapop 10.0 ,  Ygapext  0.5
               Fgapop  6.0 ,  Fgapext  7.0
               Delop   6.0 ,  Delext   7.0

Searched:      2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters:        4876514

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10054313/runat_22032004_140417_28126/runat_22032004_140417_28126/app_query.fasta_1.455
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10054313 @CGN_1_1_164 @runat_22032004_140417_28126
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  Published_Applications_NA:*
 1:  /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
 2:  /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
 3:  /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq:*
 4:  /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
 5:  /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
 6:  /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
 7:  /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
 8:  /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
 9:  /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
10:  /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
11:  /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
12:  /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
13:  /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
14:  /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
15:  /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
16:  /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
17:  /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
18:  /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

                              SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

Result No.		Score	%	Length	Pred	DB ID	Description
1		1526	98.7	1807	14	US-10-198-846-12889	Sequence 12889, A
2		1522	98.4	965	12	US-10-262-511-85	Sequence 85, Appl
3		1045.5	67.6	631	12	US-10-262-511-79	Sequence 79, Appl
4		1045.5	67.6	716	12	US-10-262-511-77	Sequence 77, Appl
5		979.5	63.4	586	12	US-10-262-511-81	Sequence 81, Appl
6	c	792	51.2	457	12	US-10-262-511-83	Sequence 83, Appl
7	c	549	35.5	310	9	US-09-796-692-9423	Sequence 9423, Ap
8	c	549	35.5	310	14	US-10-040-862-9423	Sequence 9423, Ap
9	c	549	35.5	310	15	US-10-057-475B-9423	Sequence 9423, Ap
10	c	549	35.5	310	15	US-10-154-884B-9423	Sequence 9423, Ap
11	c	541	35.0	310	9	US-09-795-692-7011	Sequence 7011, Ap
12	c	541	35.0	310	14	US-10-040-862-7011	Sequence 7011, Ap
13	c	541	35.0	310	15	US-10-057-475B-7011	Sequence 7011, Ap
14	c	541	35.0	310	15	US-10-154-884B-7011	Sequence 7011, Ap
15		468	30.3	436	10	US-09-918-995-29506	Sequence 29506, A
16		450	29.1	1279	10	US-09-814-353-19582	Sequence 19582, A
17		434	28.1	473	9	US-09-960-352-6980	Sequence 6980, Ap
18		399	25.8	407	9	US-09-960-352-3900	Sequence 3900, Ap
19		364	23.5	764	9	US-09-910-943-431	Sequence 431, App
20		338.5	21.9	910	15	US-10-369-493-36958	Sequence 36958, A
21		308	19.9	831	15	US-10-369-493-26857	Sequence 26857, A
22		276	17.9	979	15	US-10-369-493-26227	Sequence 26227, A
23		249.5	16.1	1047	15	US-10-369-493-25601	Sequence 25601, A
24		240	15.5	499	15	US-09-814-353-13698	Sequence 13698, A
25		240	15.5	501	10	US-09-814-353-943	Sequence 943, App
26		240	15.5	501	10	US-09-814-353-7313	Sequence 7313, Ap
27		236.5	15.3	462	14	US-10-260-877-71	Sequence 71, Appli
28		236.5	15.3	1830121	14	US-10-329-960-1	Sequence 1, Appli
29		235.5	15.2	1830121	15	US-10-329-670-1	Sequence 1, Appli
30		235.5	15.2	801	15	US-09-975-719-154	Sequence 154, App
31		235.5	15.2	2048	10	US-09-975-719-136	Sequence 136, App
32	c	193.5	12.5	9025608	14	US-10-156-761-1	Sequence 1, Appli
33	c	176.5	11.4	708	14	US-10-156-761-1127	Sequence 1127, App
34		171.5	11.1	1545	12	US-10-425-114-34337	Sequence 34337, A
35		158	10.2	672	15	US-10-369-493-26705	Sequence 26705, A
36		152	9.8	8332	10	US-09-006-298-1	Sequence 1, Appli
37		152	9.8	8323	14	US-09-970-597-2	Sequence 2, Appli
38		150	9.7	185548	14	US-10-175-523-62	Sequence 62, Appli
39		144	9.3	8655	14	US-10-175-523-85	Sequence 85, Appl
40	c	141.5	9.2	640681	9	US-09-790-988-1	Sequence 1, Appli
41		140.5	9.1	13049	9	US-09-231-235-1	Sequence 1, Appli
42		140.5	9.1	13049	9	US-09-797-518A-1	Sequence 1, Appli
43		140.5	9.1	13049	10	US-09-996-073-1	Sequence 1, Appli
44		140.5	9.1	13049	10	US-09-872-696A-1	Sequence 1, Appli
45		138	8.9	7984	12	US-10-359-120-162	Sequence 162, App

ALIGNMENTS

```
RESULT 1
US-10-198-846-12889
; Sequence 12889, Application US/10198846
; Publication No. US2003009974A1
; GENERAL INFORMATION:
;  APPLICANT: Lillie, James
;  APPLICANT: Xu, Yongyao
;  APPLICANT: Wang, Youzhen
;  APPLICANT: Steinmann, Kathleen
;  TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS
;  TITLE OF INVENTION: FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
;  TITLE OF INVENTION: THERAPY OF BREAST CANCER
;  FILE REFERENCE: MRI-049
;  CURRENT APPLICATION NUMBER: US/10/198,846
;  CURRENT FILING DATE: 2002-07-18
;  PRIOR APPLICATION NUMBER: 60/306,220
;  PRIOR FILING DATE: 2001-07-18
;  NUMBER OF SEQ ID NOS: 14084
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12889
;  LENGTH: 1807
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
```

```
;  NAME/KEY: misc_feature
;  LOCATION: 1802, 1803, 1804, 1805, 1806, 1807
;  OTHER INFORMATION: n = A,T,C or G
US-10-198-846-12889

Alignment Scores:
Pred. No.:              1.66e-170        Length:            1807
Score:                  1526.00          Matches:           282
Percent Similarity:     99.30%           Conservative:      2
Best Local Similarity:  98.60%           Mismatches:        2
Query Match:            98.71%           Indels:            0
DB:                                      Gaps:              0

US-10-054-313-1 (1-286) x US-10-198-846-12889  (1-1807)

Qy    1  MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg    20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  112  ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCCTTGCCCTCGCCGCCCGC   171

Qy   21  GlySerArgArgGlyPheGlyMetPheTyrAlaValArgGlyLysThrGlyValPhe    40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  172  GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGCCGCAAGAACCGGGGTCTTT   231

Qy   41  LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys    60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  232  CTGACCTGGAATGAGTGCAGAGCACAGTGGACCGGTTCCTGCTGCCAGATTTAAGAAG   291

Qy   61  PheAlaThrGluAspGluAlaThrTrpAlaPheValArgLysSerAlaSerProGluValSer    80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  292  TTTGCCACAGAGGATGAGGCCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA   351

Qy   81  GluGlyHisGluAsnGlnHisGlyGlyGluSerGluAlaLysProGlyLysArgLeuArg    100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  352  GAAGGGCATGAGAAAATCAACATGGACAAGAAATCGGAGGCGAAAGCCAGCAAGGCGACTCCGT   411

Qy  101  GluProLeuAspAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro    120
            ||||||||||||||||||||||||||||||||||||||| :::|||||||||||||||||||
Db  412  GAGCCACTGGATGGAGATGGACAATGGAGAAAGCCAGAGCCGTATGCAAAGCACATGAAGCCG   471

Qy  121  SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal    140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  472  AGCGTGGAGCCCGCGCCCGTCTCCAGTTAGCAGAGACACGTTTCCTACATGGGAGACTTCGTC   531

Qy  141  ValValTyrThrAspGlyTyrCysCysSerSerAsnGlyArgArgProArgAlaGlyIle    160
            |||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||||
Db  532  GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGGAGCAGGAATC   591

Qy  161  GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln    180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  592  GGCGTTTACTGGGGGCCAGGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGCAG   651

Qy  181  ThrAsnGlnArgAlaGluIleHisAlaAlaAlaCysLysSerAlaIleGluGlnLysThrGln    200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  652  ACAAACCAAAGAGCGGAAATTCATGCAGCCCTGCAAAGCCATTGAACAAGCAAAGACTCAA   711

Qy  201  AsnIleAsnLysGluLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn    220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  712  AACATCAATAAACTGGTTCGTATTACACACAGTATGTTTACGATAAAATGGTATAACTAAC   771

Qy  221  TrpValGlnGlnTyrTrpLysLysAsnGlyTrpLysTrpSerArgAlaGlyValIleAsn    240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  772  TGGGTTCAAGGTTGGAAGAAAAATGGTGGGAAGACAAGTGCAGGGAAAGAGTGATCAAC   831

Qy  241  LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis    260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  832  AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT   891

Qy  261  ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly    280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  892  GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA   951

Qy  281  AlaLysGlnSerGlnGluAsp    286
            ||||||||||||
```

Db   952 GCTAAACAATCGGAGGAC   969

RESULT 2
US-10-262-511-85
; Sequence 85, Application US/10262511
; Publication No. US20040038223A1
; GENERAL INFORMATION:
; APPLICANT: Smithson, Glennda
; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John A.
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Ju, Jingfang
; APPLICANT: Li, Li
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Paturajan, Meera
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Ort, Tatiana
; APPLICANT: Gorman, Linda
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Anderson, David W.
; APPLICANT: Zhong, Mei
; APPLICANT: Catterton, Elina
; APPLICANT: Ji, Weizhen
; APPLICANT: Miller, Charles E.
; APPLICANT: Rastelli, Luca
; APPLICANT: Stone, David J.
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Shinkets, Richard A.
; APPLICANT: Rothenberg, Mark E.
; APPLICANT: Leach, Martin D.
; APPLICANT: Agee, Michele L.
; APPLICANT: Berghs, Constance
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-462C
; CURRENT APPLICATION NUMBER: US/10/262,511
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/373,815
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,642
; PRIOR FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: 60/381,038
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/373,260
; PRIOR FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: 60/373,826
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/327,435
; PRIOR FILING DATE: 2001-10-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 439
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 85
; LENGTH: 965
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (10)..(867)
US-10-262-511-85

Alignment Scores:

Pred. No.:　　　　　　　　2.12e-170　　　　Length:　　　　965
Score:　　　　　　　　　　1522.00　　　　　Matches:　　　281
Percent Similarity:　　98.95%　　　　　Conservative:　2
Best Local Similarity:　98.25%　　　　Mismatches:　　3
Query Match:　　　　　　98.45%　　　　　Indels:　　　　0
DB:　　　　　　　　　　　　　　　　　　　　Gaps:　　　　　0

US-10-054-313-1 (1-286) x US-10-262-511-85 (1-965)

```
Qy      1  MetSerTrpLeuLeuPheLeuAlaHisArgValAlaAlaLeuProCysArgArg     20
Db     10  ATGAGCTGGTTCTTGTTCCTGGCCCACAGAGTCGCCTTGGCCCTGCCGCCGC         69

Qy     21  GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyLysThrGlyValPhe    40
Db     70  GGCTCTCGGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCAAGACCCGGGGTCTTT   129

Qy     41  LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys  60
Db    130  CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTCCTGCTGCCAGATTAAGAAG   189

Qy     61  PheAlaThrGluAspProValAlaTrpAlaPheValArgGlySerAlaSerProGluValSer  80
Db    190  TTTGCCACAGAGGATGAGCCCTGGGCTTTGTCAGGAAATCTGCAAGCCCGAAGTTTCA   249

Qy     81  GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysAlaLysLeuArg  100
Db    250  GAAGGGCATGAAAATCAACATGGACAAGAATCCGGAGGCGAAAGCCAGCAAGGCGACTCCGT   309

Qy    101  GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro    120
Db    310  GAGCCACTGGATGGAGATGGACATGAAAGCGCAGAGCCGTATCGCAAAGCACATGAAGCCG   369

Qy    121  SerValGluProAlaProValSerArgArgAspThrPheSerTyrMetGlyAspPhePheVal   140
Db    370  AGCGTGGAGCCCGGCCGCCTCCAGTTAGCAGAGACACGTTTCCTACATGGGAGACTTCGTC   429

Qy    141  ValValTyrThrAspGlyCysCysSerSerArgGlyArgArgLysProAlaGlyAlaGlyIle  160
Db    430  GTCGTCTACACTGATGGCTGCTGCTCCCAGTAATGGGCGTAGAAGGCCGCGAGCGGAATC   489

Qy    161  GlyValTyrTrpTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyAlaArgGln  180
Db    490  GGGCGTTACTCGGGGCCAGGCCATCCTTAAATGTAGGCATTAGACTTCCTGGGCGGCAG   549

Qy    181  ThrAsnGlnArgAlaGlyGluHisAlaAlaCysLysAlaLysAlaAlaIleGluGlnAlaLysThrGln  200
Db    550  ACAAACCAAAGAGCGGAATTCATGCAGCCTGCAAGGCCATTGAACAGCAAAGACTCAA   609

Qy    201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleThrAsn  220
Db    610  AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGTATAAACTAAC   669

Qy    221  TrpValGlnGlyTrpLysTrpLysAsnGlyTrpLysTrpSerAlaGlyLysLysValIleAsn  240
Db    670  TGGGTTCAAGGTTGGAAGTGGAAGAAAATGGGTGGAAGACAGTGCAGGGAAGAGGTGATCAAAC   729

Qy    241  LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHis  260
Db    730  AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTCAGTGCAT   789

Qy    261  ValProGlyHisSerGlyTyrPheIleGlyTyrAsnGluGluAlaAspArgLeuAlaAspArgGluGluGly  280
Db    790  GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGGAAGGA   849

Qy    281  AlaLysGlnSerGlnGluAsp  286
Db    850  GCTAAACAATCGGAAGAC    867
```

RESULT 3
US-10-262-511-79
; Sequence 79, Application US/10262511
; Publication No. US2004003823A1

Alignment Scores:
Pred. No.:　　　　　　　　3.2e-114　　　　Length:　　　　631
Score:　　　　　　　　　　1045.50　　　　Matches:　　　201
Percent Similarity:　　70.63%　　　　Conservative:　1
Best Local Similarity:　70.28%　　　　Mismatches:　　83
Query Match:　　　　　　70.63%　　　　Indels:　　　　1
DB:　　　　　　　　　　　12　　　　　　　Gaps:　　　　　1

US-10-054-313-1 (1-286) x US-10-262-511-79 (1-631)

```
Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaAlaLeuProCysArgArg 20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   14 ATGAGCTGGTTCTGTTCCTGCCCACAGAGTCGCCTTGGCCGCCTTGCCCCTGCCGCCGC 73

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgLysThrGlyValPhe 40
         |||||||||||||||||||||||||||||||||||||||||||
Db   74 GGCTTCGCGCGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCCGCAAGACCGGGGTCTTT 133

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
         |||||||||||||||||
Db  134 CTGACCTGGAATGAGTGC------- 151

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerProGluValSer 80
Db  151 ------------------------- 151

Qy   81 GluGlyHisGluAsnGlnHisHisGlyGlnGluSerGlyAlaLysProGlyLysArgLeuArg 100
Db  151 ------------------------- 151

Qy  101 GluProLeuAspGlyGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
Db  151 ------------------------- 151

Qy  121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
         ||||||||||||||||||
Db  152 -------------------AGAGACACGTTTTCCTACATGGGAGACTTCGTC 184

Qy  141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
         |||||||||||||||||||||||||||||||||:::::|||||||||||||||||||
Db  185 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCCGGAGCAGGAAATC 244

Qy  161 GlyValTyrArgGlyHisProGlyLeuAsnValGlyIleArgLeuProGlyGlyArgGln 180
         |||||||||||||||||
Db  245 GGCGTTTACTGGGGGCCGCCGGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGGCCAG 304

Qy  181 ThrAsnGlnArgAlaGlyIleHisHisAlaAlaCysLysSerAlaIleGluGlnAlaLysThrGln 200
         ||||||||||||||||||
Db  305 ACAAACCAAAGCAGGCGGAAATTCATGCAGCCTGCAAAGCCATTGACAAGCAAAGACTCAA 364

Qy  201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
         ||||||||||||||||||
Db  365 AACATCAATAAACTGGTTCTCTATACAGACAGTATGTTTACGATAAATGTATAACTAAC 424

Qy  221 TrpValGlnGlyTrpProLysLysAsnGlyTrpProLysThrSerAlaGlyValGluValIleAsn 240
         ||||||||||||||||||
Db  425 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC 484

Qy  241 LysGluAspPheValAlaAlaLeuGluArgLeuThrGlnMetAspIleGlnTrpProMetHis 260
         ||||||||||||||||||
Db  485 AAAGAGGACTTTGTGGCACTCGAGAGCCTTACCCAGGGATGGACATTCAGTGGATGCAT 544

Qy  261 ValProGlyHisSerGlyPheIleGluTyrProGlyAsnGluGlyAsnLysProArgLeuAlaAspArgLeuArgGluAlaArgGluGly 280
         ||||||||||||||||||
Db  545 GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAGAGCTGACAGATTAGCCAGAAGGA 604

Qy  281 AlaLysGlnSerGluAsp 286
         ||||||||||||||||||
Db  605 GCTAAACAATGGAAGAC 622
```

RESULT 4
US-10-262-511-77
```
; Sequence 77, Application US/10262511
; Publication No. US20040038223A1
; GENERAL INFORMATION:
;  APPLICANT: Smithson, Glenda
;  APPLICANT: Millet, Isabelle
;  APPLICANT: Peyman, John A.
;  APPLICANT: Kekuda, Ramesh
;  APPLICANT: Ju, Jingfang
```

; APPLICANT: Li, Li
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Patturajan, Meera
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Ort, Tatiana
; APPLICANT: Gorman, Linda
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Anderson, David W.
; APPLICANT: Zhong, Mei
; APPLICANT: Catterton, Elina
; APPLICANT: Ji, Weizhen
; APPLICANT: Miller, Charles E.
; APPLICANT: Rastelli, Luca
; APPLICANT: Stone, David J.
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Rothenberg, Mark E.
; APPLICANT: Leach, Martin D.
; APPLICANT: Agee, Michele L.
; APPLICANT: Berghs, Constance
```
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-462C
; CURRENT APPLICATION NUMBER: US/10/262,511
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/373,815
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,642
; PRIOR FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: 60/381,038
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/373,260
; PRIOR FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: 60/373,826
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/327,435
; PRIOR FILING DATE: 2001-10-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 439
; SOFTWARE: CurSeqList version 0.1
;
; SEQ ID NO 77
;  LENGTH: 716
;  TYPE: DNA
;  ORGANISM: Homo sapiens
; FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (10)..(618)
;
US-10-262-511-77
```

Alignment Scores:
```
Pred. No.:
Score:                          1045.50
Percent Similarity:             70.63%
Best Local Similarity:          70.28%
Query Match:                    67.63%
DB:                             12

Length:         716
Matches:        201
Conservative:   1
Mismatches:     83
Indels:         1
Gaps:           1
```

```
Pred. No.:          3.79e-114
```

US-10-054-313-1 (1-286) x US-10-262-511-77 (1-716)

```
Qy   1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaAlaLeuProCysArgArg 20
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  10 ATGAGCTGGTTCTGTTCCTGCCCACAGAGTCGCCTTGGCCGCCTTGCCCCTGCCGCCGC 69
```

APPLICANT: Malyankar, Uriel M.
APPLICANT: Ort, Tatiana
APPLICANT: Gorman, Linda
APPLICANT: Zerhusen, Bryan D.
APPLICANT: Anderson, David W.
APPLICANT: Zhong, Mei
APPLICANT: Catterton, Elina
APPLICANT: Ji, Weizhen
APPLICANT: Miller, Charles E.
APPLICANT: Rastelli, Luca
APPLICANT: Stone, David J.
APPLICANT: Pena, Carol E. A.
APPLICANT: Shenoy, Suresh G.
APPLICANT: Shimkets, Richard A.
APPLICANT: Rothberg, Mark E.
APPLICANT: Leach, Martin D.
APPLICANT: Agee, Michele L.
APPLICANT: Berghs, Constance

TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME

Alignment Scores:

Pred. No.:		
Score:	1.89e-106	Length: 586
Percent Similarity:	69.60%	Matches: 188
Best Local Similarity:	68.86%	Conservative: 0
Query Match:	63.36%	Mismatches: 83
DB:		Indels: 1
	12	Gaps:

US-10-054-313-1 (1-286) x US-10-262-511-81 (1-586)

```
Qy  14 AlaAlaAlaProCysArgArgGlySerArgGlyMetPheTyrAlaValArgArg      33
           :::|||||||||||||||||||||||||||||||||||||||||||||||
Db   8 TCCGGCCTTGCCCTGCCGCCGGGTCTTCGGGGGGTTTCATGCCGTGAGGAGG        67

Qy  34 GlyArgLysThrGlyValPheLeuThrThrAsnGluCysArgAlaGlnValAspArgPhe  53
       |||||||||||||||||||||||||||||||||||||||||
Db  68 GGCCGGAAGACCGGGGTCTTTCGACCTGGAATGAGTGC------------          106

Qy  54 ProAlaAlaArgPheLysPheAlaThrGluAlaTrpAlaPheValArgLys          73
```

RESULT 5
US-10-262-511-81
; Sequence 81, Application US/10262511
; Publication No. US20040038223A1
; GENERAL INFORMATION:

APPLICANT: Smithson, Glennda
APPLICANT: Millet, Isabelle
APPLICANT: Peyman, John A.
APPLICANT: Kekuda, Ramesh
APPLICANT: Ju, Jingfang
APPLICANT: Li, Li
APPLICANT: Guo, Xiaojia (Sasha)
APPLICANT: Patturajan, Meera
APPLICANT: Spytek, Kimberly A.
APPLICANT: Edinger, Shlomit R.
APPLICANT: Ellerman, Karen

```
Qy  21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgLysThrGlyValPhe   40
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  70 GGCTCTCGCCGGGTTCGGGGATGTTCTATGCCGTGAGGAGGGCCCAAGACCGGGGTCTTT  129

Qy  41 LeuThrThrAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys   60
       ||||||||||||||||||
Db 130 CTGACCTGGAATGAGTGC------------                                147

Qy  61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
Db 147 ------------                                                  147

Qy  81 GluGlyHisGluAsnGlnHisGluSerGluAlaLysProGlyLysArgLeuArg        100
Db 147 ------------                                                  147

Qy 101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
Db 147 ------------                                                  147

Qy 121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
                  ||||||||||||||||||||||||
Db 148 --------AGAGACCACGTTTTCCTACATGGGAGACTTCGTC                    180

Qy 141 ValValTyrThrAspCysCysSerSerAsnGlyArgArgLysProArgAlaGlyAlaGlyIle 160
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCGAGCAGGAATC  240

Qy 161 GlyValTyrTyrTrpGlyHisProLeuAsnValGlyIleArgGlyIleArgGlyArgGln  180
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GGCGTTTACTGGGGGCCCGGCCATCCTTTAAATGGCATTAGACTTCCTGGGCGGCCAG    300

Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysLysAlaIleGluGlnIleAlaLysThrGln 200
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA  360

Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC  420

Qy 221 TrpValGlnGlyTyrTrpLysLysAsnGlyTrpLysLysSerAlaGlyValLysGluValIleAsn 240
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAGAGAGTGATCAAC  480

Qy 241 LysGluAspPheValAlaLeuGluArgLeuThrThrGlnGlyMetAspIleGlnTrpMetHis 260
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGATGACATTCAGTGGATGCAT    540

Qy 261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly  280
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA  600

Qy 281 AlaLysGlnSerGluAsp                                            286
       ||||||||||||||||||
Db 601 GCTAAACAATCGGAAGAC                                            618
```

APPLICANT: Stone, David J.
; APPLICANT: Pena, Carol E. A.
; APPLICANT: Shenoy, Suresh G.
; APPLICANT: Shimkets, Richard A.
; APPLICANT: Rothenberg, Mark B.
; APPLICANT: Leach, Martin D.
; APPLICANT: Agee, Michele L.
; APPLICANT: Berghs, Constance
; TITLE OF INVENTION: NOVEL PROTEINS AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 21402-462C
; CURRENT APPLICATION NUMBER: US/10/262,511
; CURRENT FILING DATE: 2003-05-28
; PRIOR APPLICATION NUMBER: 60/326,483
; PRIOR FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: 60/373,815
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/327,917
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/381,642
; PRIOR FILING DATE: 2002-05-17
; PRIOR APPLICATION NUMBER: 60/328,029
; PRIOR FILING DATE: 2002-10-09
; PRIOR APPLICATION NUMBER: 60/381,038
; PRIOR FILING DATE: 2002-05-16
; PRIOR APPLICATION NUMBER: 60/328,056
; PRIOR FILING DATE: 2001-10-09
; PRIOR APPLICATION NUMBER: 60/373,260
; PRIOR FILING DATE: 2002-04-17
; PRIOR APPLICATION NUMBER: 60/373,826
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: 60/327,435
; PRIOR FILING DATE: 2001-10-05
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 439
; SOFTWARE: CuraSeqList version 0.1
;
; SEQ ID NO 83
; LENGTH: 457
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(457)
US-10-262-511-83

Alignment Scores:
Pred. No.:   Length: 457
Score: 2.17e-84   Matches: 145
Percent Similarity: 792.00   Conservative: 1
Best Local Similarity: 100.00%   Mismatches: 0
Query Match: 51.23%   Indels: 0
DB: 12   Gaps: 0

US-10-054-313-1 (1-286) x US-10-262-511-83 (1-457)

US-10-054-313-1 (1-286) x US-10-262-511-83 (1-457)

Qy   137 GlyAspPheValValValTyrThrAspGlyCysCysSerAsnGlyArgArgLysPro 156
          ||||||||||||||||||||||||||||||||||| :::||
Db   11 GGAGACTTCGTCGTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCG 70

Qy   157 ArgAlaGlyIleGlyValTyrValTyrProGlyHisProLeuAsnValGlyIleArgLeu 176
          |||||||||||||||||||||||||||||||||||||||||||
Db   71 CGAGCAGGAATCGGCGTTTACTGGGGCCGGGCCATCCTTTAAATGTAGGCATTAGACTT 130

Qy   177 ProGlyArgGlnThrArgGlnArgAlaGlnAlaIleHisAlaAlaCysLysAlaIleGluGln 196
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   131 CCTGGGCGGCAGACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAA 190

Qy   197 AlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsn 216
          ||||||||||||||||||||||||||||||||||||||
Db   191 GCAAAGACTCAAAACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAAT 250

Qy   217 GlyIleThrAsnTrpValGlnTrpValGlyAsnGlyTrpLysLysTrpThrSerAlaGlyLys 236
          ||||||||||||||||||||||||||||||||||||||||
Db   251 GGTATAACTAACTGGTTCAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAA 310

Db   106 -- 106
Qy   74 SerAlaSerProGluValSerGluHisGluAsnGlnHisGlyGlnGluSerGluAla 93
Db   106 -- 106
Qy   94 LysProGlyLysArgLeuArgGluArgGluLeuAspGlyHisGlySerAlaGlnPro 113
Db   106 -- 106
Qy   114 TyrAlaLysHisMetLysProSerValGluProAlaProProValSerArgAspThrPhe 133
Db   107 ------------------------------------AGAGACACGTTT 118
Qy   134 SerTyrMetGlyAspPheValValValTyrThrAspGlyCysCysSerSerAsnGlyArg 153
          ||||||||||||||||||||||||||||||||||||
Db   119 TCCTACATGGGAGACTTCGTCGTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGT 178
Qy   154 ArgLysProGlyArgAlaGlyIleGlyValTyrValTyrProGlyHisProLeuAsnValGly 173
          |||::::||||||||||||||||||||||||||||||||||||||
Db   179 AGAAGGCCGCGAGCAGGAATCGGCGTTTACTGGGGCCGGGCCGAGCCATCCTTTAAATGTAGGC 238
Qy   174 IleArgLeuProGlyArgGlnThrArgGlnArgAlaGlnAlaIleHisAlaAlaCysLysAla 193
          ||||||||||||||||||||||||||||||||||||||||||||
Db   239 ATTAGACTTCCTGGGCGGCAGACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCC 298
Qy   194 IleGluGlnAlaLysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPhe 213
          |||||||||||||||||||||||||||||||||||||
Db   299 ATTGAACAAGCAAAGACTCAAAACATCAATAAACTGGTTCTGTATACAGACAGTATGTTT 358
Qy   214 ThrIleAsnGlyIleThrAsnTrpValGlnTrpValGlyAsnGlyTrpLysThrSer 233
          |||||||||||||||||||||||||||||||||||
Db   359 ACGATAAATGGTATAACTAACTGGGTTGTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGT 418
Qy   234 AlaGlyLysGluValIleAsnLysAspPheValAlaLeuGluLeuThrGlnGly 253
          |||||||||||||||||||||||||||||||||||||
Db   419 GCAGGGAAAGAAGAGGTGATCAACAAAGAGGACTTTGTGCACTGGAGAGGCTTACCCAGGGGG 478
Qy   254 MetAspIleGlnTrpPheMetHisValProLeuHisSerGlyPheIleGlyAsnGluGluAla 273
          |||||||||||||||||||||||||||||||||||||
Db   479 ATGGACATTCAGTGGATGATGCATGTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCT 538
Qy   274 AspArgLeuLeuArgGluGluGluArgGluLysAlaLysGlyAlaLysLysGlnSerGluAsp 286
          |||||||||||||||||||||||||
Db   539 GACAGATTAGCCAGAGAGAGAGAGCTAAACAATGGGAAGAC 577

RESULT 6
US-10-262-511-83, Application US/10262511
; Sequence 83, Application No. US20040038223A1
; Publication No. US20040038223A1
; GENERAL INFORMATION:
; APPLICANT: Smithson, Glenda
; APPLICANT: Millet, Isabelle
; APPLICANT: Peyman, John A.
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Ju, Jingfang
; APPLICANT: Li, Li
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Patturajan, Meera
; APPLICANT: Spytek, Kimberly A.
; APPLICANT: Edinger, Shlomit R.
; APPLICANT: Ellerman, Karen
; APPLICANT: Malyankar, Uriel M.
; APPLICANT: Ort, Tatiana
; APPLICANT: Gorman, Linda
; APPLICANT: Zerhusen, Bryan D.
; APPLICANT: Anderson, David W.
; APPLICANT: Zhong, Mei
; APPLICANT: Catterton, Elina
; APPLICANT: Ji, Weizhen
; APPLICANT: Miller, Charles E.
; APPLICANT: Rastelli, Luca

```
Qy   237  GluValIleAsnLysGluAspPheValAlaLeuGluArgThrGlnGlyMetAspIle  256
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   311  GAGGTGATCAACAAGAGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATT  370

Qy   257  GlnTrpMetHisValProGlyHisSerGlyPheIleGlyPheAsnGluAlaAspArgLeu  276
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   371  CAGTGGATGCATGTTCCGTGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTA  430

Qy   277  AlaArgGluGlyAlaLys  282
          ||||||||||
Db   431  GCCAGAGAAGGAGCTAAA  448
```

RESULT 7
US-09-796-692-9423/c
; Sequence 9423, Application US/09796692
; Publication No. US2002019836A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9423
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-9423

Alignment Scores:
Pred. No.:   7.67e-56   Length: 310
Score: 549.00   Matches: 98
Percent Similarity: 98.04%   Conservative: 2
Best Local Similarity: 96.08%   Mismatches: 2
Query Match: 35.51%   Indels: 0
DB: Gaps: 0

US-10-054-313-1 (1-286) x US-09-796-692-9423 (1-310)

```
Qy    68  TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis  87
          |||||||||||||||||||||||||||||||||||||||:::|||||||
Db   308  TGGGCCTTTGTCAGGAAAATCTGCAAGCCCCGAAGTTCAGAAGGGCATGGAAAATCAACAT  249

Qy    88  GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluProLeuAspGlyAspGly  107
```


```
Db   248  GGACAAGAATCGGAGCGAGGGCAAAGCCAGCAGCCACTCCGTGCCTGAGCCACTGGAGATCGGATCGGAGAGGA  189

Qy   108  HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro  127
          |||||||:::|||||||
Db   188  CATGAAAGCGCAGAGCCGTATGCAAAGCACATGACACATGACACATGACACATG...  129

Qy   128  ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCys  147
          ||||||||||||||||||||||||||||||||||||
Db   128  GTTAGCAGAGACACGTTTCCACATGGAGACTTCGTCGTCGTCTACACTGATGGCTGC  69

Qy   148  CysSerSerAsnGlyArgArgLysProArgAlaGlyIleGlyValTyrTrpGlyProGly  167
          |||||||||||||||:::|||||
Db    68  TGCTCCAGTAATGGGCTAGAAGGCCCGCGACGCAGGAATCGGCGTTTACTGGGGCCAGGC  9

Qy   168  HisPro  169
          ||||||
Db     8  CATCCT  3
```

RESULT 8
US-10-040-862-9423/c
; Sequence 9423, Application US/10040862
; Publication No. US2003007839GA1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Thera
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9423
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-9423

Alignment Scores:
Pred. No.:   7.67e-56   Length: 310
Score: 549.00   Matches: 98
Percent Similarity: 98.04%   Conservative: 2
Best Local Similarity: 96.08%   Mismatches: 2
```

;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-057-475B-9423

| Query Match: | 35.51% | Indels: | 0 | | Alignment Scores: | | Length: | 310 |
| DB: | | Gaps: | 0 | | Pred. No.: | 7.67e-56 | Matches: | 98 |
| | | | | | Score: | 549.00 | Conservative: | 2 |
| | | | | | Percent Similarity: | 98.04% | Mismatches: | 2 |
| | | | | | Best Local Similarity: | 96.08% | Indels: | 0 |
| | | | | | Query Match: | 35.51% | Gaps: | 0 |
| | | | | | DB: | | | |

US-10-054-313-1 (1-286) x US-10-040-862-9423 (1-310)

US-10-054-313-1 (1-286) x US-10-057-475B-9423 (1-310)

```
Qy    68  TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis  87
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   308  TGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCAGGAAGGGCATGAAAATCAACAT 249
```

```
Qy    68  TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis  87
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   308  TGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCAGGAAGGGCATGAAAATCAACAT 249
```

```
Qy    88  GlyGlnGluSerGluAlaLysProGlyLysArgLeuArgGluArgGluProLeuAspGlyAspGly 107
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   248  GGACAAGAATCGGAGGCGAGGCGAAAGCCAGCAAGCCACTCCGTGAGCCACTGGATGGAGATGGA 189
```

```
Qy   108  HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127
          ||||||||:||:||||||||||||||:|||||||||||||||||||||||||||||||||
Db   188  CATGAAAGCGCAGAGCCGTATGCAAAGCCAGCCGAGCCGTGGAGCCGGCCGCCGCCCTCCA 129
```

```
Qy   128  ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyCys 147
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   128  GTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTCGTCGTCTACACTGATGGCTGC  69
```

```
Qy   148  CysSerSerAsnGlyArgArgLysProArgAlaGlyLysIleGlyValTyrTrpGlyProGly 167
          |||||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||
Db    68  TGCTCCAGTAATGGGCGTAGAAGGCCGCGGCAGGAATCGGCGTTTACTGGGGGGCCAGGC   9
```

```
Qy   168  HisPro 169
          ||||||
Db     8  CATCCT  3
```

RESULT 9
US-10-057-475B-9423/c
;   Sequence 9423, Application US/10057475B
;   Publication No. US2004000068A1
;   GENERAL INFORMATION:
;   APPLICANT: Gaiger, Alexander
;   APPLICANT: Algate, Paul A.
;   APPLICANT: Mannion, Jane
;   APPLICANT: Clapper, Jonathan David
;   APPLICANT: Wang, Aijun
;   APPLICANT: Ordonez, Nadia
;   APPLICANT: Carter, Lauren
;   APPLICANT: McNeill, Patricia Dianne
;   APPLICANT: Corixa Corporation
;   TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;   TITLE OF INVENTION: Hematological Malignancies
;   FILE REFERENCE: 014058-014402US
;   CURRENT APPLICATION NUMBER: US/10/057,475B
;   CURRENT FILING DATE: 2002-01-22
;   PRIOR APPLICATION NUMBER: US 60/186,126
;   PRIOR FILING DATE: 2000-03-01
;   PRIOR FILING DATE: 2000-03-01
;   PRIOR FILING DATE: 2000-03-17
;   PRIOR APPLICATION NUMBER: US 60/190,479
;   PRIOR APPLICATION NUMBER: US 60/200,545
;   PRIOR FILING DATE: 2000-04-27
;   PRIOR FILING DATE: 2000-04-28
;   PRIOR APPLICATION NUMBER: US 60/200,303
;   PRIOR APPLICATION NUMBER: US 60/200,779
;   PRIOR FILING DATE: 2000-04-28
;   PRIOR APPLICATION NUMBER: US 60/200,999
;   PRIOR FILING DATE: 2000-05-04
;   PRIOR FILING DATE: 2000-05-01
;   PRIOR APPLICATION NUMBER: US 60/202,084
;   PRIOR FILING DATE: 2000-05-04
;   PRIOR APPLICATION NUMBER: US 60/206,201
;   PRIOR FILING DATE: 2000-05-22
;   PRIOR FILING DATE: 2000-07-14
;   PRIOR APPLICATION NUMBER: US 60/218,950
;   PRIOR APPLICATION NUMBER: US 60/222,903
;   PRIOR FILING DATE: 2000-08-03
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 10979
;   SOFTWARE: FastSEQ for Windows Version 3.0
;   SEQ ID NO 9423
;   LENGTH: 310

RESULT 10
US-10-154-884B-9423/c
;   Sequence 9423, Application US/10154884B
;   Publication No. US20040005561A1
;   GENERAL INFORMATION:
;   APPLICANT: Gaiger, Alexander
;   APPLICANT: Algate, Paul A.
;   APPLICANT: Mannion, Jane
;   APPLICANT: Retter, Marc W.
;   APPLICANT: Corixa Corporation
;   TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Thera
;   TITLE OF INVENTION: Hematological Malignancies
;   FILE REFERENCE: 014058-01352US
;   CURRENT APPLICATION NUMBER: US/10/154,884B
;   CURRENT FILING DATE: 2002-05-23
;   PRIOR APPLICATION NUMBER: US 60/186,126
;   PRIOR FILING DATE: 2000-03-01
;   PRIOR FILING DATE: 2000-03-17
;   PRIOR APPLICATION NUMBER: US 60/190,479
;   PRIOR APPLICATION NUMBER: US 60/200,545
;   PRIOR FILING DATE: 2000-04-27
;   PRIOR APPLICATION NUMBER: US 60/200,303
;   PRIOR FILING DATE: 2000-04-28
;   PRIOR FILING DATE: 2000-04-28
;   PRIOR APPLICATION NUMBER: US 60/200,779
;   PRIOR APPLICATION NUMBER: US 60/200,999
;   PRIOR FILING DATE: 2000-05-01
;   PRIOR APPLICATION NUMBER: US 60/202,084
;   PRIOR FILING DATE: 2000-05-04
;   PRIOR APPLICATION NUMBER: US 60/206,201
;   PRIOR FILING DATE: 2000-05-22
;   PRIOR APPLICATION NUMBER: US 60/218,950
;   PRIOR FILING DATE: 2000-07-14
;   PRIOR APPLICATION NUMBER: US 60/222,903
;   PRIOR FILING DATE: 2000-08-03

```
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 9423
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-154-884B-9423

Alignment Scores:
Pred. No.:               7.67e-56     Length:         310
Score:                   549.00       Matches:        98
Percent Similarity:      98.04%       Conservative:   2
Best Local Similarity:   96.08%       Mismatches:     2
Query Match:             35.51%       Indels:         0
DB:                                   Gaps:           0

US-10-054-313-1 (1-286) x US-10-154-884B-9423 (1-310)

Qy   68  TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis  87
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  308  TGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCAGAAGGGCATGAAAATCAACAT 249

Qy   88  GlyGlnGluSerGluAlaLysProGlyLysProGluArgLeuArgGluProLeuAspGlyGly 107
         ||||||||||||||||||||||||||||||||||||||||||||||
Db  248  GGACAAGAATCGGAGGGGCGAAAGCCAGGCGACTCCGTGAGCCACTGGATGCGATGAGGATGA 189

Qy  108  HisGluSerAlaGlnProTyrAlaAlaLysHisMetLysProSerValGluProAlaProPro 127
         ||||||||||||||::::||:|||||||||
Db  188  CATGAAAGCCGAGACCGTATGCAAAGCACATGAAAGCCCGAGCGTGGAGCCGCCGCCTCCA 129

Qy  128  ValSerArgArgAspThrPheSerTyrMetGlyValValTyrThrAspGlyCys 147
         |||||||||||||||||||||||||||||
Db  128  GTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTCGTCGTCTACACTGATGGCTGC 69

Qy  148  CysSerSerAsnGlyArgArgLysProArgGlyAlaGlyIleGlyValTyrTrpGlyTyrProGly 167
         |||||||||||::::||:|||||||||
Db   68  TGCTCCAGTAATGGGCCGTAGAAGGCCGCGAGCAGGAGAATCGGCGTTTACTGGGGGGCCAGGC 9

Qy  168  HisPro 169
         ||||||
Db    8  CATCCT 3
```

```
RESULT 11
US-09-796-692-7011/c
; Sequence 7011, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT:  Gaiger, Alexander
; APPLICANT:  Algate, Paul A.
; APPLICANT:  Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
```

```
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7011
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-7011

Alignment Scores:
Pred. No.:               6.79e-55     Length:         310
Score:                   541.00       Matches:        97
Percent Similarity:      97.06%       Conservative:   2
Best Local Similarity:   95.10%       Mismatches:     3
Query Match:             34.99%       Indels:         0
DB:                                   Gaps:           0

US-10-054-313-1 (1-286) x US-09-796-692-7011 (1-310)

Qy   68  TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis  87
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  308  TGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCAGAAGGGCATGAAAATCAACAT 249

Qy   88  GlyGlnGluSerGluAlaLysProGlyLysProGluArgLeuArgGluProLeuAspGly 107
         |||||||||||||||||||||||||||||||||||||||||||||
Db  248  GGACAAGAATCGGAGGGGCGAAAGCCAGGCGACTCCGTGAGCCACTGGATGCGATGAGGATGA 189

Qy  108  HisGluSerAlaGlnProTyrAlaAlaLysHisMetLysProSerValGluProAlaProPro 127
         ||||||||||::::||:|||||||||
Db  188  CATGAAAGCCGAGACCGTATGCAAAGCACATGAAAGCCCGAGCGTGGAGCCGCCGCCTCCA 129

Qy  128  ValSerArgArgAspThrPheSerTyrMetGlyValValTyrThrAspGlyCys 147
         |||||||||||||||||||||||||
Db  128  GTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTCGTCGTCTACACTGATGGCTGC 69

Qy  148  CysSerSerAsnGlyArgArgLysProArgGlyAlaGlyIleGlyValTyrTrpGlyTyrProGly 167
         |||||||||::::||:|||||||||
Db   68  TGCTCCAGTAATGGGCCGTAGAAGGCCGCGAGCAGGAGAATCGGCGTTTACTGGGGGCCGGGC 9

Qy  168  HisPro 169
         ||||||
Db    8  CATCCT 3
```

```
RESULT 12
US-10-040-862-7011/c
; Sequence 7011, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT:  Gaiger, Alexander
; APPLICANT:  Algate, Paul A.
; APPLICANT:  Mannion, Jane
; APPLICANT:  Retter, Marc
; APPLICANT:  Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Thera
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-01352US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
```

; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: US 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: US 60/223,378
; PRIOR FILING DATE: 2000-08-07
; PRIOR APPLICATION NUMBER: US 09/796,692
; PRIOR FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 10467
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7011
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-040-862-7011

Alignment Scores:
Pred. No.:   6.79e-55   Length: 310
Score: 541.00   Matches: 97
Percent Similarity: 97.06%   Conservative: 2
Best Local Similarity: 95.10%   Mismatches: 3
Query Match: 34.99%   Indels: 0
DB: 14   Gaps: 0

US-10-054-313-1 (1-286) x US-10-040-862-7011 (1-310)

Qy 68 TrpAlaPheValArgLysSerAlaSerGluSerGluValSerGluGlyHisGluAsnGlnHis 87
Db 308 TGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCAGAGGGCATGAAGAAAATCAACAT 249

Qy 88 GlyGlnGluSerGluAlaLysProGlyLyArgLeuArgGluProLeuAspGlyAspGly 107
Db 248 GGACAAGAATCGGAGGCGAAAGCCAAGCAGCGACTCCGTGAGCCACTGGATGGAGATGGA 189

Qy 108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127
Db 188 CATGAAAGCGCAGAGCCGTATGCAAAGCACATGAAGTCGAGCGTGGAGCCGGCGCCTCCA 129

Qy 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValTyrThrAspGlyGlyCys 147
Db 128 GTTAGCAGAGACACGTTTCTACATGGGAGACTTCGTCGTCGTTACACTGATGGCTGC 69

Qy 148 CysSerSerAsnGlyArgArgLysProArgArgAlaGlyIleGlyValTyrTrpGlyProGly 167
Db 68 TGCTCCAGTAATGGGCGTAGAAGGCCCGCAGGAATCGGCGTTTACTGGGGGGCCGGGC 9

Qy 168 HisPro 169
Db   8 CATCCT 3

RESULT 13
US-10-057-475B-7011/c
; Sequence 7011, Application US/10057475B
; Publication No. US2004002068A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Clapper, Jonathan David
; APPLICANT: Wang, Aijun
; APPLICANT: Ordonez, Nadia
; APPLICANT: Carter, Lauren
; APPLICANT: McNeill, Patricia Dianne
; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Thera
; TITLE OF INVENTION: Hematological Malignancies
; FILE REFERENCE: 014058-01440US
; CURRENT APPLICATION NUMBER: US/10/057,475B
; CURRENT FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 60/222,903
; PRIOR FILING DATE: 2000-08-03
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 10979
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7011
; LENGTH: 310
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-057-475B-7011

Alignment Scores:
Pred. No.:   6.79e-55   Length: 310
Score: 541.00   Matches: 97
Percent Similarity: 97.06%   Conservative: 2
Best Local Similarity: 95.10%   Mismatches: 3
Query Match: 34.99%   Indels: 0
DB: 15   Gaps: 0

US-10-054-313-1 (1-286) x US-10-057-475B-7011 (1-310)

Qy 68 TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis 87
Db 308 TGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCAGAGGGCATGAAGAAAATCAACAT 249

Qy 88 GlyGlnGluSerGluAlaLysProGlyLyArgLeuArgGluProLeuAspGlyAspGly 107
Db 248 GGACAAGAATCGGAGGCGAAAGCCAAGCAGCGACTCCGTGAGCCACTGGATGGAGATGGA 189

Qy 108 HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro 127
Db 188 CATGAAAGCGCAGAGCCGTATGCAAAGCACATGAAGTCGAGCGTGGAGCCGGCGCCTCCA 129

Qy 128 ValSerArgAspThrPheSerTyrMetGlyAspPheValValTyrThrAspGlyGlyCys 147
Db 128 GTTAGCAGAGACACGTTTCTACATGGGAGACTTCGTCGTCGTTACACTGATGGCTGC 69

Qy 148 CysSerSerAsnGlyArgArgLysProArgArgAlaGlyIleGlyValTyrTrpGlyProGly 167
Db 68 TGCTCCAGTAATGGGCGTAGAAGGCCCGCAGGAATCGGCGTTTACTGGGGGGCCGGGC 9

Qy 168 HisPro 169
Db   8 CATCCT 3

RESULT 14
US-10-154-884B-7011/c
; Sequence 7011, Application US/10154884B
; Publication No. US20040005561A1
; GENERAL INFORMATION:

```
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc W.
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy
;   of Hematological Malignancies
; FILE REFERENCE: 014058-013521US
; CURRENT APPLICATION NUMBER: US/10/154,884B
; CURRENT FILING DATE:   2002-05-23
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE:  2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE:  2000-04-27
; PRIOR FILING DATE:  2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE:  2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE:  2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE:  2000-05-01
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE:  2000-05-04
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE:  2000-05-22
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE:  2000-07-14
; PRIOR APPLICATION NUMBER: US 60/218,950
; PRIOR FILING DATE:  2000-08-03
; PRIOR APPLICATION NUMBER: US 60/222,903
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 11290
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 7011
;   LENGTH: 310
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-154-884B-7011

Alignment Scores:
Pred. No.:
Score:                    541.00           Length:        310
Percent Similarity:       97.06%           Matches:       97
Best Local Similarity:    95.10%           Conservative:  2
Query Match:              34.99%           Mismatches:    3
DB:                                        Indels:        0
                          6.79e-55         Gaps:          0

US-10-054-313-1 (1-286) x US-10-154-884B-7011 (1-310)

Qy      68  TrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGlnHis  87
                ||||||||||||||||||||||||||||||||||:||||||||||
Db     308  TGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAGTTCGAAGAGGCATGAAAATCAACAT  249

Qy      88  GlyGlnGluSerGluAlaLysProGlyTyrLysProGlyLeuArgGluProLeuAspGlyAspGly  107
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     248  GGACAAGAATCGGAGGCGAGGCGAAAGCCAGCAGCGACTCCGTGAGCGCCACTGACTGGAGATGGA  189

Qy     108  HisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaProPro  127
                ||||||||||||||:::||||||||||||||||||||||||||||||||||
Db     188  CATGAAAGCCCAGAGCCCGTATGCAAAGCACATGAAGTCCAGCCGTGGAGCCGCCCTCCA  129

Qy     128  ValSerArgAspThrPheSerTyrMetGlyAspPheValValValTyrThrAspGlyGlyCys  147
                ||||||||||||||||||||||||||||||||||||||||||||||
Db     128  GTTAGCAGAGACACAGTTTTCCTACATGGGAGACTTCGTCGTCTACACTGACTGGCTGC  69

Qy     148  CysSerSerAsnGlyArgArgLysProArgArgAlaGlyTyrIleGlyTyrTyrThrGlyTyrProGlyGly  167
                ||||||||||:::||||||||||||||||||||||||||||||||||||||||||||:
Db      68  TGCTCCAGTAATGGGCGTAGAAAGGCCGCCGGAATCGGCGTTACTGGGGGCCGGGGC  9

Qy     168  HisPro  169
                ||||||
Db       8  CATCCT  3
```

```
RESULT 15
US-09-918-995-29506
; Sequence 29506, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION:   FROM VARIOUS cDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE:  2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 29506
;   LENGTH: 436
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(436)
;   OTHER INFORMATION: n = A, T, C or G
US-09-918-995-29506

Alignment Scores:
Pred. No.:                                 Length:        436
Score:                    468.00           Matches:       86
Percent Similarity:       100.00%          Conservative:  3
Best Local Similarity:    96.63%           Mismatches:    0
Query Match:              30.27%           Indels:        0
DB:                       4.74e-46         Gaps:          0

US-10-054-313-1 (1-286) x US-09-918-995-29506 (1-436)

Qy     198  LysThrGlnAsnIleAsnLysAsnLeuValLeuTyrThrAspSerMetPheThrIleAsnGly  217
                :::||||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db      32  CAGACTCAAAACATCAATAAAACTGGTTCTGTATACAGACAGTATGTTTACTTTAAATGGT  91

Qy     218  IleThrAsnTrpValGlnGlyTyrTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGlu  237
                ||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      92  ATTTCTAACTGGGTTCAAGGTTACTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAG  151

Qy     238  ValIleAsnLysGluAspPheValAlaLeuArgLeuArgThrGlnGlyMetAspIleGln  257
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     152  GTGATCAACAAAGAGACTTTGTGGCACTGGAGAGGCTTACCCAGGGATGGACATTCAG  211

Qy     258  TrpMetHisValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAla  277
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     212  TGGATGCATGTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATAGCC  271

Qy     278  ArgGluGlyAlaLysGlnSerGluAsp  286
                |||||||||||||||||||||
Db     272  AGAGAAGGAGCTAAACAATCGGAAGAC  298

Search completed: March 23, 2004, 00:17:10
Job time : 303 secs
```

OM protein - nucleic search, using frame_plus_p2n model

Copyright GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: March 22, 2004, 22:30:37 ; Search time 2335 Seconds
(without alignments)
3657.638 Million cell updates/sec

Title: US-10-054-313-1
Perfect score: 1546
Sequence: 1 MSWLLPLAHRVALALPCRR.........FIGNEEADRLAREGAKQSED 286

Scoring table: BLOSUM62
                Xgapop 10.0 , Xgapext 0.5
                Ygapop 10.0 , Ygapext 0.5
                Fgapop 6.0 , Fgapext 7.0
                Delop   6.0 , Delext   7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10054313/runat_22032004_140416_28082/app_query.fasta_1.455
-DB=EST -QFMT=fastp -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10054313 @CGN_1_1_2135 @runat_22032004_140416_28082 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
            1: em_estba:*
            2: em_esthum:*
            3: em_estin:*
            4: em_estmu:*
            5: em_estov:*
            6: em_estpl:*
            7: em_estro:*
            8: em_htc:*
            9: gb_esti:*
           10: gb_est2:*
           11: gb_htc:*
           12: gb_est3:*
           13: gb_est5:*
           14: gb_est4:*
           15: em_estfun:*
           16: em_estom:*
           17: em_gss_hum:*
           18: em_gss_inv:*
           19: em_gss_pln:*
           20: em_gss_vrt:*
           21: em_gss_fun:*
           22: em_gss_mam:*
           23: em_gss_mus:*
           24: em_gss_pro:*
           25: em_gss_rod:*
           26: em_gss_phg:*
           27: em_gss_vrl:*
           28: gb_gss1:*
           29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1501	97.1	1201	13	BX394121	BX394121 BX394121
2	1498	96.9	1201	9	AL560874	AL560874 AL560874
3	1455	94.1	1054	12	BM541420	BM541420 AGENCOURT
4	1446	93.5	1060	12	BM810651	BM810651 AGENCOURT
5	1432	92.6	952	9	AL554334	AL554334 AL554334
6	1415.5	91.6	944	12	BG829086	BG829086 60275428
7	1406	90.9	1196	13	BX416456	BX416456 BX416456
8	1345	87.0	893	13	BU175613	BU175613 AGENCOURT
9	1340	86.7	949	12	BG831321	BG831321 60276160
10	1288	83.3	765	14	CA775004	CA775004 ip11g06.y
11	1281	82.9	935	13	BQ921568	BQ921568 AGENCOURT
12	1239.5	80.2	779	10	BE746490	BE746490 601579445
13	1229	79.5	818	13	BU595053	BU595053 AGENCOURT
14	1222	79.0	916	12	BM451915	BM451915 AGENCOURT
15	1215	78.6	869	10	BF984400	BF984400 602307923
16	1210	78.3	707	12	BI223765	BI223765 602943149
17	1200	77.6	959	10	BF689462	BF689462 602186827
18	1198.5	77.5	714	10	BE778327	BE778327 601463720
19	1195	77.3	946	10	BF038506	BF038506 601460421
20	1155	74.7	764	12	BG767598	BG767598 602741757
21	1153	74.6	1198	12	BM805842	BM805842 AGENCOURT
22	1110.5	71.8	693	12	BG768406	BG768406 602742157
23	1106.5	71.6	828	29	AY404958	AY404958 Homo sapi
24	1100	71.2	864	12	BG420594	BG420594 602448473
25	1090.5	70.5	782	13	BU929741	BU929741 AGENCOURT
26	1089.5	70.5	772	14	CB961021	CB961021 AGENCOURT
27	1083.5	70.1	792	10	BE407918	BE407918 601299260
28	1067	69.0	613	12	BM782115	BM782115 K-EST0058
29	1058	68.4	652	12	BI518784	BI518784 603061972
30	1058	68.4	686	12	BG824728	BG824728 602728713
31	1052.5	68.1	937	10	BE786259	BE786259 601474394
32	1040.5	67.3	704	10	BF129039	BF129039 60181367
33	1018	65.8	1209	13	BM454170	BM454170 AGENCOURT
34	1003.5	64.9	868	13	BQ942688	BQ942688 AGENCOURT
c	35	1002	64.8	783	14	CB529180	CB529180 UI-H-FT2-
36	996	64.4	1364	11	AK011680	AK011680 Mus muscu
37	992	64.2	927	12	BI558492	BI558492 603240375
38	973	62.9	823	13	BQ896738	BQ896738 AGENCOURT
39	968.5	62.6	1102	12	BI410112	BI410112 602964039
40	967	62.5	619	12	BG719686	BG719686 602689896
41	964	62.4	857	14	CF617505	CF617505 AGENCOURT
c	42	960.5	62.1	718	12	BG778946	BG778946 602667481
c	43	952	61.6	801	10	BF690400	BF690400 602186827
44	949	61.4	534	12	BM839989	BM839989 K-EST0116
c	45	945	61.1	807	9	AW006811	AW006811 wt07f11.x

ALIGNMENTS

RESULT 1
BX394121
LOCUS   BX394121   1201 bp   mRNA   linear   EST 13-MAY-2003
DEFINITION BX394121 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
            cDNA clone CSODC010YI23 5-PRIME, mRNA sequence.
ACCESSION   BX394121
VERSION   BX394121.1 GI:30612409
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1201)

AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   Contact: Genoscope
           Genoscope - Centre National de Sequencage
           BP 191 91006 EVRY cedex - France
           Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
           Library was constructed by Life Technologies, a division of
           Invitrogen. This sequence belongs to sequence cluster 1105.r For
           more information about this cluster, see
           http://www.genoscope.cns.fr/
           cgi-bin/cluster.cgi?seq=CS0DC010AE12QP1&cluster=1105.r. Contact :
           Feng Liang Email : fliang@lifetech.com URL :
           http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
           Faraday Avenue Genoscope sequence ID : CS0DC010AE12QP1.
FEATURES   Location/Qualifiers
  source   1..1201
           /organism="Homo sapiens"
           /mol_type="mRNA"
           /db_xref="taxon:9606"
           /clone="CS0DC010YJ23"
           /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
           /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
           /note="1st strand cDNA was primed with a NotI-oligo(dT)
           primer. Five prime end enriched, double-strand cDNA was
           digested with Not I and cloned into the Not I and EcoR V
           sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Alignment Scores:
Pred. No.:   1.87e-130
Score: 1501.00
Percent Similarity: 99.30%
Best Local Similarity: 98.25%
Query Match: 97.09%
DB: 13

US-10-054-313-1 (1-286) x BX394121 (1-1201)

```
Qy    1 MetSerTrpLeuPheLeuAlaHisArgValAlaAlaLeuAlaAlaLeuProCysArgArg   20
          :::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   99 YTGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCTTGCCCTGCCGCGC  158

Qy   21 GlySerArgGlyPheGlyMetPheTyrValAlaValArgGlyValArgGlyValThrGlyValPhe   40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  159 GGCTCTCGCGGGTTCGGAATGTTCTATGCCGTGAGGAGGGCCGCAAGAGACCGGGGTCTTT  218

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAlaAspArgPheProAlaAlaArgPheLysLys   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  219 CTGACCCGGAATGAGTGCAGAGCACAGGTGGACCGGTTCCCTGCTGCCAGATTTAAGAAG  278

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  279 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAAATCTGCAAGCCCGGAAGTTTCA  338

Qy   81 GluGlyHisGluAsnGlnHisGlyGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  339 GAAGGAGCATGAAAATCAACATGGACAAGAATCGGAGGCGAAAGT-AGCAAGCGACTCCGT  397

Qy  101 GluProLeuArgGlyAspGlyHisGluSerArgAlaGlnProTyrAlaLysHisMetLysPro  120
          |||||||||||||||||||||||||||||:|||||||||||||||||||||||||||
Db  398 GAGCCACTGGATCGAGATGACAATGGACATGAAAGCCGAGAGCCGTATGCAAAGCACATGAAGCCG  457

Qy  121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  458 AGCGTGGAGCCCGCCCCAGTTAGCAGACGACACGTTTTCCTACATGGGAGACTTCGTC  517

Qy  141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyAlaGlyIle  160
          |||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db  518 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAGGCCGGAGCAGGAAATC  577

Qy  161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyTyrIleArgGlyProGlyArgGln  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db 578 GGGGTTTACTGGGGGCCAGGCCATCCTTAAATGTAGGCATTAGACTTCCTCGGCGCGGGCAG 637

Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaAlaIleGluGlnIleAlaLysThrGln 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 638 ACAAACCAAGAGAGCGGAAATTCATGCAGCCTGCAGCCTGCAAAGCCTCGAACAACAAGCAAAGAACTCAA 697

Qy 201 AsnIleAsnLysLeuValLeuTyrThrArgSerMetPheThrIleAsnGlyIleThrAsn 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 698 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAAACTAAC 757

Qy 221 TrpValGlnGlyTyrTrpLysLysAsnGlyTyrTrpSerThrSerAlaGlyTyrGluValIleAsn 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 758 TGGGTTCAAGGTTCGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAAC 817

Qy 241 LysGluAspPheValAlaLeuGluThrGlnHisGlyMetAspIleGlnTrpMetHis 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 818 AAAGAGGACTTTGTGGCACTGGAGGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 877

Qy 261 ValProGlyHisSerGlyPheIleGlyValGluAlaArgLeuAlaArgGlyLeuGlyGly 280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 878 GTTCCTGGTVCATTCGGGATTTATAGGCAATGAGAGTGACAGAGATTAGCCAGAGAAGGA 937

Qy 281 AlaLysGlnSerGluAsp 286
          ||||||||||||||||
Db 938 GCTAAACAATCGGAAGAC 955

RESULT 2
AL560874
LOCUS   AL560874   1201 bp   mRNA   linear   EST 31-MAY-2003
DEFINITION AL560874 Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED
            Homo sapiens cDNA clone CS0DL005YJ14 5-PRIME, mRNA sequence.
ACCESSION   AL560874
VERSION   AL560874.2 GI:31285003
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE   Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
  COMMENT   On Feb 15, 2001 this sequence version replaced gi:12907756.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 1105.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DL005DB07QP1&cluster=1105.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ invitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DL005DB07QP1.
FEATURES   Location/Qualifiers
  source   1..1201
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /clone="CS0DL005YJ14"
            /cell_type="B CELLS (RAMOS CELL LINE)"
            /cell_line="RAMOS CELL LINE"
            /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE) COT
            25-NORMALIZED"
            /note="1st strand cDNA was primed with a NotI-oligo(dT)
            primer. Five prime end enriched, double-strand cDNA was
            digested with Not I and cloned into the Not I and EcoR V
            sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN   Length: 1201

Alignment Scores:
Pred. No.:   3.58-130

```
Score:                    1498.00     Matches:        281
Percent Similarity:       98.95%      Conservative:     2
Best Local Similarity:    98.25%      Mismatches:       3
Query Match:              96.90%      Indels:           1
DB:                                   Gaps:             0

US-10-054-313-1 (1-286) x AL560874 (1-1201)

Qy     1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaAlaLeuAlaAlaLeuProCysArgArg 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   142 ATGAGCTGGCTTCGTTCCTGGCCCACAGAGTCGCCGCTGCCCTTGCCCTGCCCGCCGC 201

Qy    21 GlySerArgArgPheGlyMetPheTyrAlaValArgArgGlyLysThrGlyValPhe 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   202 GGCTTCCGCGGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 261

Qy    41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   262 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG 321

Qy    61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   322 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 381

Qy    81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   382 GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGGCAAAGC-AGCAGGACTCCGT 440

Qy   101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
           |||||||||||||||| |:||||||||||||||||||||||||||||||||||||||
Db   441 GAGCCACTGGATGGAGGATGGACATGAAAGCGCAGAGCCCTATGCAAAGCACATGAAGCCG 500

Qy   121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   501 AGCGTGGAGCCGGCCGCCCTCAGTAGCAGACACGTTTCCTACATGGGAGACTTCGTC 560

Qy   141 ValValTyrThrAspGlyCysSerSerArgAsnGlyArgArgLysProArgArgAlaGlyIle 160
           |||||||||||||||||||||||||||||||||:::||||||||||||||||||||||
Db   561 GTCGTCTACACTGATGGCTGCTCGTCCCAGTAATGGGCGTAGAAGGCCGGAGCAGGAATC 620

Qy   161 GlyValTyrTrpProGlyHisProLeuAsnValGlyIleGlyIleArgLeuProGlyArgGln 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   621 GGCGTTTACTGGGGCCAGGCCATCCTTTAAATGTAGGCATTAGGACTTCCTGGGCGGCAG 680

Qy   181 ThrAsnGlnArgAlaGlyIleHisIleAlaAlaCysLysLysAlaIleGluGlnHisAlaLysThrGln 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   681 ACAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCCAAGCCAATTGAACACAGCAAAGACACTCAA 740

Qy   201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   741 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGTATAACTAAC 800

Qy   221 TrpValGlnGlyGlnGluSerGluAlaGlyLysThrLysSerAlaGlyLysGluValIleAsn 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   801 TGGGTTCAAGGTTGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAGAGGTGATCAAC 860

Qy   241 LysGluAspPheValAlaAlaLeuAlaLeuGluArgGlnMetAspIleGlnTrpMetHis 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   861 AAAGAGGACTTTGTGCCACTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTCAGTGCAT 920

Qy   261 ValProGlyHisSerGlyPheIleGlyLeuGlyAsnGluGluAlaAspArgLeuAlaArgGluGly 280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   921 GTTCCTGGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCACAGAAGAAGGA 980

Qy   281 AlaLysGlnSerGluAsp 286
           ||||||||||||||||||
Db   981 GCTAAACAATCGGAAGAC 998
```

```
RESULT 3
BM541420
LOCUS        BM541420      1054 bp    mRNA        linear    EST 20-FEB-2002
DEFINITION   AGENCOURT_6493820 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5521285
             5', mRNA sequence.
ACCESSION    BM541420
VERSION      BM541420.1  GI:18770049
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 1054)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM12186 row: k  column: 14
             High quality sequence stop: 713.
FEATURES             Location/Qualifiers
     source          1..1054
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5521285"
                     /tissue_type="leiomyosarcoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_71"
                     /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.1 kb. "
ORIGIN
```

```
Alignment Scores:
Pred. No.:                3.24e-126    Length:        1054
Score:                    1455.00      Matches:        278
Percent Similarity:       97.24%       Conservative:     4
Best Local Similarity:    95.86%       Mismatches:       4
Query Match:              94.11%       Indels:           4
DB:                                    Gaps:             0

US-10-054-313-1 (1-286) x BM541420 (1-1054)

Qy     1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaAlaAlaLeuAlaAlaLeuProCysArgArg 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    45 ATGAGCTGGCTTCGTTCCTGGCCCACAGAGTCGCCTTGGCCGCCTTGCCCTGCCCGCCGC 104

Qy    21 GlySerArgArgPheGlyMetPheTyrAlaValArgArgGlyLysThrGlyValPhe 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   105 GGCTTCCGCGGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 164

Qy    41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   165 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG 224

Qy    61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   225 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 284

Qy    81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   285 GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGGCAAAGCGAGCAGGACTCCGT 344

Qy   101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   345 GAGCCACTGGATGGAGATGGACATGAAAGCGCAGAGCCCTATGCAAAGCACATGAAGCCG 404

Qy   121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   405 AGCGTGGAGCCGGCCGCCCTCAGTAGCAGACACGTTTCCTACATGGGAGACTTCGTC 464
```

```
                                        in the laboratory of Gerald M. Rubin (University of
                                        California, Berkeley) using ZAP-cDNA synthesis kit
                                        (Stratagene) and Superscript II RT (Life Technologies).
                                        Note: this is a NIH_MGC Library."

ORIGIN

Alignment Scores:
Pred. No.:          2.3e-125       Length:              1060
Score:              1446.00        Matches:              274
Percent Similarity: 96.86%         Conservative:           4
Best Local Similarity: 95.47%      Mismatches:             8
Query Match:        93.53%         Indels:                 2
DB:                 12             Gaps:                   0

US-10-054-313-1 (1-286) x BM810651 (1-1060)

Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg   20
          |||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db   58 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCTTTGGCCGCCCTGCCCCTGCCGCGC  117

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgLysThrGlyLysValValPhe   40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  118 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGCCCAAGACCGGGGGTCTT  177

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  178 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG  237

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  238 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA  297

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysLysArgLeuArg  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  298 GAAGGGCATGAAAATCAACATGACAAGAAATCGGAGCAAGAAAGCCAGCAGACTCCGT  357

Qy  101 GluProLeuAspGlyPheGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro  120
          |||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||
Db  358 GAGCCACTGGATGGAGATGGACATGAGTCAGCCCTATGCAAAGCATATGAAGCCG  417

Qy  121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  418 AGCGTGGAGCCGGCGCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC  477

Qy  141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle  160
          |||||||||||||||||:||:|||||||||||||||||||||||||||||||||||||||
Db  478 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGCGGCAGGAAGCCGAGCAGGAATC  537

Qy  161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleValArgGlyArgGln  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  538 GGCGTTTACTGGGGCCCGGGCCATCCTTAAATGTAGGCATTAGAACTTCCTGGGCGCAG  597

Qy  181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGluAsnGlyThrGln  200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  598 ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA  657

Qy  201 AsnIleAsnLysAsnLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  658 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGTATAACTAAC  717

Qy  221 TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsn  240
          |||||||||||||||||||||||:::
Db  718 TGGGTTCAAGGTTGGAAGAA-AATGGTGAAGACAAGTGCAGGAAAGAGGTGATNCAC  776

Qy  241 LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleLeuTrpMetHis  260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  777 AAAGAGGACTTTGTGCACTGGAGGCTTACCCAGGGGATGGACATTCAGTGGATGCAT  836

Qy  261 ValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArg-LeuAlaArgGluGlyI  280
          ||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db  837 GGTTCCTGGTCATTCGGGATTTATANGGCATGAAGAAGCTGCAGATTTANCCAGGAAGG  896

Qy  280 yAlaLysGlnSerGluAsp  286
```

(continued on left portion)

```
Qy  141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle  160
          |||||||||||||||||:||:|||:|||||||||||||||||||||||||||||||||||
Db  465 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGGCAGGAAGCCGAGCAGGAATC  524

Qy  161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleValArgGlyArgGln  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  525 GGCGTTTACTGGGGCCCGGGCCATCCTTAAATGTAGGCATTAGAACTTCCTGGGCGGCAG  584

Qy  181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGluAsnGlyThrGln  200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  585 ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA  644

Qy  201 AsnIleAsnLysAsnLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  645 AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC  704

Qy  221 TrpValGlnGlyTrpLys-LysAsnGlyTrpLysThrSerAla-GlyLysGluValIleA  240
          ||||||||||||||||||:||||||||||||||||||||:
Db  705 TGGGTTCAAGGTTGGAAGNAAAAATGGTGAAGACAAGTGCAGGAAAGAGGTGATCA  764

Qy  240 snLysGluAspPheValAlaLeuGluArgLeuThrGln-GlyMetAspIleGlnTrpMet  259
          |||||||||||||||||||||||||||||||||:
Db  765 ACAAAGAGGACTTTGTGCACTGGAGGCTTACCCAGGGGNATGGACATTCAGTGGATG  824

Qy  260 HisValPro-GlyHisSerGlyPheIleGlyGluGluAlaAspArgLeuAlaArgG1  279
          |||||||||||||||||||||||||||||:||:
Db  825 CATGTTCCTGGTCATTCGGGATTTATAGGCCATGAAGAAACTGACACAGATTAGCCAGGAA  884

Qy  279 uGlyAlaLysGlnSerGluAsp  286
          |||||||||||||||||||
Db  885 AGGAGCTAAACCATCCGGAAGAC  906
```

```
RESULT 4
BM810651
LOCUS       BM810651              1060 bp    mRNA       linear    EST 05-MAR-2002
DEFINITION  AGENCOURT_6580549 NIH_MGC_98 Homo sapiens cDNA clone IMAGE:5453776
            5', mRNA sequence.
ACCESSION   BM810651
VERSION     BM810651.1  GI:19127474
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1060)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1947  row: n  column: 17
            High quality sequence stop: 725.
FEATURES            Location/Qualifiers
     source         1..1060
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:5453776"
                    /tissue_type="astrocytoma grade IV, cell line"
                    /lab_host="DH10B (phage-resistant)"
                    /clone_lib="NIH_MGC_98"
                    /note="Organ: brain; Vector: pOTB7; Site 1: XhoI; Site 2:
                    EcoRI; cDNA made by oligo-dT priming. Directionally
                    cloned into EcoRI/XhoI sites using the following 5'
                    adaptor: GGCACCAG(G). Library constructed by Ling Hong
```

Db   897 AGCTNAACAATCGGAAGAC 915
          |||| ||||||||||||||||
Db   392 GAAGGGCATGAAATCAACATGGACAAGAATCGGAGGCGAAGCCAGCCAGCGACTCCGT 451

RESULT 5
AL554334
LOCUS   AL554334   952 bp   mRNA   linear   EST 31-MAY-2003
DEFINITION   AL554334 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
             clone CS0DI082YL10 5-PRIME, mRNA sequence.
ACCESSION   AL554334
VERSION   AL554334.2   GI:31276146
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 952)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   On Feb 15, 2001 this sequence version replaced gi:12895013.
             Contact: Genoscope
             Genoscope - Centre National de Sequencage
             BP 191 91006 EVRY cedex - France
             Email: seqref@genoscope.cns.fr; Web : www.genoscope.cns.fr
             Library was constructed by Life Technologies, a division of
             Invitrogen. This sequence belongs to sequence cluster 1105.r For
             more information about this cluster, see
             http://www.genoscope.cns.fr/
             cgi-bin/cluster.cgi?seq=CS0DI082DF05QP1&cluster=1105.r. Contact :
             Feng Liang Email : fliang@lifetech.com URL :
             http://fullength.invitrogen.com/ Invitrogen Corporation 1600
             Faraday Avenue Genoscope sequence ID : CS0DI082DF05QP1.
FEATURES   Location/Qualifiers
     source   1..952
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /clone="CS0DI082YL10"
                     /db_xref="taxon:9606"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Alignment Scores:
Pred. No.:   Length: 952
Score: 4.04e-124   Matches: 263
Percent Similarity: 99.25%   Conservative: 2
Best Local Similarity: 98.50%   Mismatches: 2
Query Match: 92.63%   Indels: 0
DB: Gaps: 0

US-10-054-313-1 (1-286) x AL554334 (1-952)

Qy   1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 152 ATGAGCTGGCTTCTTCTTCTTCGGCCACAGAGTCGCCTTGGCCGTCCCTTGCCCTGCCGCGC 211

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyLysThrGlyValPhe 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 212 GGCTTCTCGGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 271

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 272 CTGACCTGGAATGAGTGCAGAGCAGAGCAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG 331

Qy   61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 332 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 391

Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGlnGluLysProGlyLysProGluLeuArg 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy 101 GluProLeuAspGlyAspGlyHisGlySerAlaGlnLeuProTyrAlaLysHisMetLysPro 120
          |||||||||||||||||||||||||||||||||||||||||||: :::||||||||||||
Db 452 GAGCCACTGGATGGAGATGGACATGAGAAAGCCAGAGCCGATGCCAAGCACATGAAGCCG 511

Qy 121 SerValGluProAlaProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 512 AGCGTGGAGCCCGCCCTCCAGTTAGCAGGACACGTTTCCTACATGGGAGACTTCGTC 571

Qy 141 ValValTyrThrAspGlyCysSerSerAsnGlyArgArgLysProArgAlaGlyIle 160
          |||||||||||||||||||||||||||||||||||||||||||::::|||||||||
Db 572 GTCGTCTACACTGATGGCTGCTCCTCCAGTAATGGGCGTAGAGGGCCGGAGCAGGAATC 631

Qy 161 GlyValTyrTrpGlyHisProLeuAsnValGlyIleArgArgLeuProGlyArgGln 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 632 GGCGTTTACTGGGGGCCAGGCCATCTTTAAATGAGCATTAGACTTCCTGGGGCGCAG 691

Qy 181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaAlaIleGluGlnAlaLysThrGln 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 692 ACAAACCAAAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA 751

Qy 201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 752 AACATCAATAAACTCGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC 811

Qy 221 TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysValIleAsn 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 812 TGGGTTCAAGGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGTGATCAAC 871

Qy 241 LysGluAspPheValAlaLeuGluThrGlnGlyMetAspIleGlnTrpMetHis 260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 872 AAAGAGGACTTTGTGGCACTGGAGACGCTTACCCAGGGGATGGACATTCAGTGGATGCAT 931

Qy 261 ValProGlyHisSerGlyPhe 267
          |||||||||||||||||
Db 932 GTTCCTGGTCATTCGGGATTT 952

RESULT 6
BG829086
LOCUS   BG829086   944 bp   mRNA   linear   EST 22-MAY-2001
DEFINITION   602752428F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4905246 5',
             mRNA sequence.
ACCESSION   BG829086
VERSION   BG829086.1   GI:14176673
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 944)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: ATCC
             cDNA Library Preparation: Ling Hong/Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLCM1806 row: g column: 07
             High quality sequence stop: 781.
FEATURES   Location/Qualifiers
     source   1..944
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4905246"
                     /tissue_type="rhabdomyosarcoma"

```
                                                  Qy  260 isValProGlyHisSerGlyPheIleGlyAlaGluGluAlaAspArgLeuAlaArgGluG 280
                                                           |||||||||           :::      ::|||||||||||||||||||||||
                                                  Db  784 ATGTACCGGGTCATCGGGATTAT---AGGCAATGAAGAAGCTGCACAGATTAGCCAGAGAAG 840

                                                  Qy  280 lyAlaLysGlnSerGluAsp 286
                                                           |||||            ||:
                                                  Db  841 GAGTTAA-CAATCGGAAGAC 859


                                                  RESULT 7
                                                  LOCUS        BX416456             1196 bp      mRNA      linear    EST 15-MAY-2003
                                                  DEFINITION   BX416456 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
                                                               CS0DA007YO12 5-PRIME, mRNA sequence.
                                                  ACCESSION    BX416456
                                                  VERSION      BX416456.1  GI:30763629
                                                  KEYWORDS     EST.
                                                  SOURCE       Homo sapiens (human)
                                                  ORGANISM     Homo sapiens
                                                               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                                                               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
                                                  REFERENCE    1  (bases 1 to 1196)
                                                  AUTHORS      Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
                                                  TITLE        Full-length cDNA libraries and normalization
                                                  JOURNAL      Unpublished (2001)
                                                  COMMENT      Contact: Genoscope
                                                               Genoscope - Centre National de Sequencage
                                                               BP 191 91006 EVRY cedex - France
                                                               Email: seqref@genoscope.cns.fr; Web : www.genoscope.cns.fr
                                                               Library was constructed by Life Technologies, a division of
                                                               Invitrogen. This sequence belongs to sequence cluster 1105.r For
                                                               more information about this cluster, see
                                                               http://www.genoscope.cns.fr/
                                                               cgi-bin/cluster.cgi?seq=CS0DA007BH06Q1&cluster=1105.r. Contact :
                                                               Feng Liang Email : fliang@lifetech.com URL :
                                                               http://fullength.invitrogen.com/ InVitroGen Corporation 1600
                                                               Faraday Avenue Genoscope sequence ID : CS0DA007BH06QP1.
                                                  FEATURES             Location/Qualifiers
                                                  source       1..1196
                                                               /organism="Homo sapiens"
                                                               /mol_type="mRNA"
                                                               /db_xref="taxon:9606"
                                                               /clone="CS0DA007YO12"
                                                               /tissue_type="NEUROBLASTOMA"
                                                               /clone_lib="Homo sapiens NEUROBLASTOMA"
                                                               /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                                                               with a NotI-oligo(dT) primer. Five prime end enriched,
                                                               double-strand cDNA was digested with Not I and cloned into
                                                               the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                                                               Library was not normalized."
                                                  ORIGIN

                                                  Alignment Scores:
                                                  Pred. No.:                        Length:       1196
                                                  Score:           1.61e-121        Matches:       277
                                                  Percent Similarity:     97.55%     Conservative:    2
                                                  Best Local Similarity:  96.85%     Mismatches:      7
                                                  Query Match:            90.94%     Indels:          4
                                                  DB:                    13          Gaps:            0

                                                  US-10-054-313-1 (1-286) x BX416456  (1-1196)

                                                  Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaAlaAlaAlaLeuAlaProCysArgArg 20
                                                           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                  Db  125 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCGCCGCCGCCCTTGGCCCGTGCGCCGC 184

                                                  Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValAlaArgGlyArgLysThrGlyValPhe 40
                                                           |||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                  Db  185 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 244

                                                  Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAlaAspArgPheAlaAlaArgPheLysLys 60
                                                           |||     |||:::||||||||||:|  ||  |:|||||:||:|
                                                  Db  245 CTGACCTGGAATGAGTGCAGAGCCCAGGTGGCACAGGTGGACCGGTTTCCTGCTGCCAGATTAAGAAG 304
```

```
                                                  ORIGIN

                                                  Alignment Scores:
                                                  Pred. No.:                        Length:        944
                                                  Score:           1.42e-122        Matches:       271
                                                  Percent Similarity:     96.18%     Conservative:    6
                                                  Best Local Similarity:  94.10%     Mismatches:      8
                                                  Query Match:            91.56%     Indels:          4
                                                  DB:                    12          Gaps:            1

                                                  US-10-054-313-1 (1-286) x BG829086  (1-944)

                                                  Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaAlaAlaAlaLeuAlaProCysArgArg 20
                                                           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                  Db    4 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCTTGCCCTGCCCGTGCGCCGC 63

                                                  Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValAlaArgGlyArgLysThrGlyValPhe 40
                                                           |||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                  Db   64 GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 123

                                                  Qy   41 LeuThrTrpAsnGluCysArgAlaGlnValAlaAspArgPheAlaAlaArgPheLysLys 60
                                                           |||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                  Db  124 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTAAGAAG 183

                                                  Qy   61 PheAlaThrGluAspGluAlaGlnAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
                                                           |||:|||||||:|:|||:||||:||||||||||||||||||||||||||||||||
                                                  Db  184 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 243

                                                  Qy   81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyGlyAlaLysArgLeuArg 100
                                                           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                  Db  244 GAAGGGCATGAAAATCAACATGCAAGAAATCGGAGCCAAGAAGCCAGCAAGCGACTCCGT 303

                                                  Qy  101 GluProLeuAspGlyHisGlySerAlaGlnProTyrArgAlaLysHisMetLysPro 120
                                                           |||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                  Db  304 GAGCCACTGGATGGACATGGGCAGCCATCTTAAATGGCATTAGGCAGCCATGAAGCCG 363

                                                  Qy  121 SerValGluProAlaProProValSerArgSerArgAspThrProSerTyrMetGlyAspPheVal 140
                                                           |||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                  Db  364 AGCGTGGAGCCGGCCGCCTCCAGTAGCAGAGCACGTTTTCCTACATGGGAGACTTCGTC 423

                                                  Qy  141 ValValTyrThrAspGlyCysCysSerSerSerAsnGlyArgArgLysProArgArgAlaGlyIle 160
                                                           |||||||||||||||||||||||||||||||||::||||||||||||||||||||
                                                  Db  424 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGCCGGAGGCAGGAAATC 483

                                                  Qy  161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyValIleArgLeuProGlyArgGln 180
                                                           |||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                  Db  484 GGCGTTTACTGGGGGCCAGGCCATCTTAAATGGCGTCGTTTACGATAAATGGTATAACTAA 543

                                                  Qy  181 ThrAsnGlnArgAlaGluIleHisAlaAlaCys-LysAlaIleGluGlnAlaLysThrGl 200
                                                           |||||||||||||||||||||||||||||||||||||||||||||||||||||
                                                  Db  544 ACAAACCAAAGAGCGAAATTCATGCAGCCTGCAAAAGCCATTGAACACAGCAAAGACTCA 603

                                                  Qy  200 nAsnIleAsnLysArgLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAs 220
                                                           |||||||||||||||||||||||||||||||||||||||||||||||||::||||
                                                  Db  604 AAACATCAATAAACTGTTCGTATACAGACAGTATGTTTACGATAAATGGTATAACTAA 663

                                                  Qy  220 nTrpValGlnGlyTrpLysLysAsnGlyTrpLysTrpSerAlaGlyLysLysGluValIleAs 240
                                                           |||||||||||||||||||||||||||||||||||||||:::
                                                  Db  664 CTGGGTTCAAGGTTGGAAGAAAATGGTGGAAGAAAGTGCAGGCAAAGAGGTGATCAA 723

                                                  Qy  240 nLysGluAspPheValAla-LeuGluArgLeuThrGlyMetAspIleGlnTrpMetH 260
                                                           ||:||||||||:||:|:||  ||::|||||||||||||||||||||||
                                                  Db  724 CAAAGAGGACTTGGTGGCACTGGAGAGGCTTACCAAGGGATGGACATTCAGTGGATGC 783
```

                                                  ```
                                                  /lab_host="DH10B (phage-resistant)"
                                                  /clone_lib="NIH_MGC_17"
                                                  /note="Organ: muscle; Vector: pOTB7; Site 1: EcoRI;
                                                  Site 2: XhoI; cDNA made by oligo-dT priming.
                                                  Directionally cloned into EcoRI/XhoI sites using the
                                                  following 5' adaptor: GGCACGAG(G).  Size-selected >500bp
                                                  for average insert size 1.8kb.  Library constructed by
                                                  Ling Hong in the laboratory of Gerald M. Rubin (University
                                                  of California, Berkeley) using ZAP-cDNA synthesis kit
                                                  (Stratagene) and Superscript II RT (Life Technologies)."
                                                  ```

```
                                          http://image.llnl.gov
                                          Place: LLAM13196 row: h column: 24
                                          High quality sequence stop: 672.
                          FEATURES                Location/Qualifiers
                          source            1..893
                                          /organism="Homo sapiens"
                                          /mol_type="mRNA"
                                          /db_xref="taxon:9606"
                                          /clone="IMAGE:6009671"
                                          /tissue_type="large cell carcinoma"
                                          /lab_host="DH10B (phage-resistant)"
                                          /clone_lib="NIH_MGC_68"
                                          /note="Organ: lung; Vector: pCMV-SPORT6, Site 1: NotI;
                                          Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                                          Average insert size 1.8 kb.  Library constructed by Life
                                          Technologies. "

                          ORIGIN
   Alignment Scores:
   Pred. No.:                             893
   Score:              5.61e-116          Length:          893
   Percent Similarity: 1345.00            Matches:         270
   Best Local Similarity: 95.14%          Conservative:      4
   Query Match:        93.75%             Mismatches:        4
   DB:                 87.00%             Indels:           10
                       13                 Gaps:              0

   US-10-054-313-1 (1-286) x BU175613 (1-893)

   Qy     1 MetSerTrpLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg      20
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db    27 ATGAGCTGGCTTCTGTTCCTGCCCCACAGAGTCGCCCTTGCCGCTCTGCCGTGTCGT      86

   Qy    21 GlySerArgGlyMetPheTyrAlaValArgArgGlyThrGlyValPhe      40
            ||||||||||||||||||||||||||||||||||||||||||||||||||
   Db    87 GGCTCTCGCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT      146

   Qy    41 LeuThrTrpAsnGluCysArgAlaGlnValAspArgPheProAlaAlaArgPheLysLys      60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   147 CTGACCTGGAATGAGTGCAGAGCAGGTGCAGCCGTTTCCTGCTGCCAGATTTAAGAAG      206

   Qy    61 PheAlaAlaThrGluAspGluAlaTrpAlaArgLysSerAlaSerProGluValSer      80
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   207 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAAATCTGCAAGCCCGGAAGTTTCA      266

   Qy    81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg     100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   267 GAAGGGCATGAAAATCAACATGGACAAGGCCGGAAAGCCGAAAGCCAAGCAAGCGACTCCGT     326

   Qy   101 GluProLeuAspGlyArgGlyHisAspGlyHisGluSerAlaGlnInProTyrAlaLysHisMetLysProPro     120
            |||||||||||||||||||||||||||||||||||||||:::::||||||||||||||||||
   Db   327 GAGCCACTGGATGGACGATGGACATGAAAAGCCCAGAGCCGTATGCAAAGCACATGAAGCCG     386

   Qy   121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal     140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   387 AGCGTGGAGCCCGCGCCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC      446

   Qy   141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgProArgAlaGlyIle     160
            |||||||||||||||||||||||||||||||||::|||||||||||||||||||||||
   Db   447 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCAGCAGGAATC      506

   Qy   161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln     180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   507 GGCGTTTACTGGGGGCCGGGGCCATCCTTTAAATGTAGGCATTAGGACTTCCTGGGGCCGCCAG     566

   Qy   181 ThrArgAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln     200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   567 ACAAACCAAAGAGCGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA      626

   Qy   201 AsnIleAsnLysValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn     220
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   627 AACATCAATAAACTGGTTCTGTATACAGATAGTATGTTTACGATAAATGGTATAACTAAC      686

   Qy   221 TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleValIleAsn     240
```

(Second column, left side of page)

```
   Qy    61 PheAlaThrGluAspGluAlaTrpAlaArgLysSerAlaSerProGluValSer      80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   305 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAAATCTGCAAGCCCGGAAGTTTCA      364

   Qy    81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg     100
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   365 GAAGGGCATGAAAATCAACATGGACAAGGCCGGAAAGC-AGCMAGGCGACTCCGT      423

   Qy   101 GluProLeuAspGlyArgGlyHisAspGlyHisGlnProTyrAlaLysHisMetLysPro     120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   424 GAGCCACTGGATGGACGATGGACATGAAAAGCCCAGAGCCGTATGCAAAGCACATGAAGCCG      483

   Qy   121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal     140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   484 AGCGTGGAGCCCGCGCCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC      543

   Qy   141 ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgProArgAlaGlyIle     160
            |||||||||||||||||||||||||||||||||::||||||||||||||||||
   Db   544 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCAGCAGGAATC      603

   Qy   161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln     180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   604 GGGGTTTACTGGGGGCCAGGCCATCCTTTAAATGTAGGCATTAGGACTTCCTGGGGCCGCCAG      663

   Qy   181 ThrArgAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln     200
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   664 AMAAACCAAAGAGCGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA      723

   Qy   201 AsnIleAsnLysValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn     220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   724 AACATCAATAAACTGGTTCTGTATACAGATAGTATGTTTACGATAAATGGTATAACTAAC      783

   Qy   221 TrpValGlnGlyTrpLysAsnGlyTrpLysThrSerAlaGlyLysGluValIleValIleAsn     240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   784 TGGGTTCAAGGTTGGAAGAA-AATGGGGTGGAAGACAAGTGCAGGGAAAGAGTGATCAC-      841

   Qy   241 LysGluAspPheValAlaAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpProMetHis     260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   842 AAAGAGGACTTTGTGGCACTGGAGAGGCTTACCCAGGGGATGGACATTCAGTGGCATGCAT      901

   Qy   261 ValProGlyHisGlyPheIleGlyPheIleGluGlyAsnGluGluAlaAspArgArgGluGluGly     280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
   Db   902 GTYCT-GGTCATTCGGGATTTATAGGCAATGAAGAAGCTGACAGATTAGCCAGAGAAGGA      960

   Qy   281 AlaLysGlnSerGluAsp     286
            ||||||||||||||||||
   Db   961 GCTAAACAATCGGAAGAC     978
```

```
RESULT 8
BU175613
LOCUS       BU175613                 893 bp    mRNA    linear    EST 04-SEP-2002
DEFINITION  AGENCOURT 7935024 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:6009671
            5', mRNA sequence.
ACCESSION   BU175613
VERSION     BU175613.1  GI:22689597
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 893)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
```

```
Db   687 TGGGTTCAAGGTTGGAAGAAAATGGTGAGGAAGAAGTGCAGGGAAGAGAGTGATCAAC 746
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy   241 LysGluAspPhe-ValAlaLeu-GluArgLeuThrGlnGly--MetAspIleGlnTrp-M 259
             ||||||| |  |||  |  |||   ||| | |  |:::||||  |  |||  | ||
Db   747 AAAGAGGACTTTGGTGCACTGGAGAAGCTTACCCAGGGGNATGGGACATTCAGTGGGGA 806

Qy   259 etHis-ValProGly-His-SerGlyPheIle-GlyAsn-GluGluAlaAspArgLeuAl 277
             ||| || ||||| ||| |||||||||| ||||| |||||||||||||| ||||||
Db   807 TGCATGGTTCCTGNCCATTCGGGATTTATAGGGCAATGGAAGAAACTGACAGATTAGC 866

Qy   277 aArg 278
         |:::
Db   867 CCAG 870
```

RESULT 9
BG831321
LOCUS   BG831321   949 bp   mRNA   linear EST 22-MAY-2001
DEFINITION 602766160F1 NIH_MGC_42 Homo sapiens cDNA clone IMAGE:4908326 5',
            mRNA sequence.
ACCESSION   BG831321
VERSION   BG831321.1 GI:14178908
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 949)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM1814 row: g column: 15
            High quality sequence stop: 807.
FEATURES   Location/Qualifiers
     source   1..949
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4908326"
                     /tissue_type="epithelioid carcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_42"
                     /note="Organ: pancreas; Vector: pOTB7; Site 1: XhoI;
                     Site 2: EcoRI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                     for average insert size 1.8kb. Library constructed by Ling
                     Hong in the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

Alignment Scores:
Pred. No.:   1.82e-115   Length: 949
Score: 1340.00   Matches: 257
Percent Similarity: 93.93%   Conservative: 6
Best Local Similarity: 91.79%   Mismatches: 15
Query Match: 86.68%   Indels: 3
DB: 12   Gaps: 0

US-10-054-313-1 (1-286) x BG831321 (1-949)

Qy   1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaAlaLeuProCysArgArg 20
```

```
Db    30 ATGAGCTGGCTTCGTTCCTGGCCCACAGAGTCGGCTTGGCGCCGCCTTGCCCGCCTGCCGGCC 89
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Qy    21 GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyArgLysThrGlyValPhe 40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    90 GGCTCTCGCGGGTTCCGGGTGTTCTATGCCGTGAGGAGGGGCCAAGACCGGGGTCTTT 149

Qy    41 LeuThrTrpAsnGluCysArgAlaGlnValAspAlaArgPheProAlaAlaArgPheLysLys 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   150 CTGACCTGGAATGAGTGCAGAGCACAGGTGACCGGTTCCTGCGTTGCCAGATTTAAGAAG 209

Qy    61 PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   210 TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCGGAAGTTTCA 269

Qy    81 GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaAlaProGlyLysArgLeuAsn 100
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   270 GAAGGGCATGAAAATCAACATGGACAAGAATCGGAGGCAGCAGCCAGGGACTCCGT 329

Qy   101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
         |||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db   330 GAGCCACTGGATGGAGATGGACATGAAAGGCCAGAGCCGTATGCAAAGCACATGAAGCCG 389

Qy   121 SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal 140
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   390 AGCGTGGAGCCGGCGCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC 449

Qy   141 ValValThrThrAspGlyCysSerSerAsnGlyArgArgLysProArgArgAlaGlyIle 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   450 GTCGTCTCACACTGGTTGCTGCTCCAGTAATGGGCGTAGAAGGCCGGAGCCAGGAATC 509

Qy   161 GlyValTyrTrpGlyProGlyHisProLeuAsnValGlyValIleArgLeuProGlyArgGln 180
         |||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db   510 GGGCGTTACTGGGGACCCGGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGGCAG 569

Qy   181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln 200
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   570 ACAAACCAAAGAGCGGCAGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA 629

Qy   201 AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleLysGlyIleThrAsn 220
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db   630 AACATCAATAAACTGGTTCTGTATCACAGACCAGTATGTTTACGATAAAATGGTATAAACTAAC 689

Qy   221 TrpValGlnGlyTyrTrp-LysLysAsn-GlyTyrTrpLysThrSerAlaGlyLysGluValIleA 240
         ||||||||||||||||||  |||||||  |  ||| ||   |||:::
Db   690 TGGGTTCAAGGTTGGAAGAAAATGGTTGGAAGACAAGTGCAGGGCAAAGAGGTGATC 749

Qy   240 snLysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetH 260
             ||||||| |  |||  |  |||   ||| | |  |:::||||  |  |||  | ||
Db   750 AAAAAGAGGACTTTGTGGGACTGGAGAGGTT-ACCAGGGGATGGGACATTCAGTGGGATGC 808

Qy   260 isValProGlyHisSerGlyPheIleGlyAsnGluGluAlaAspArgLeuAlaArg 278
         |||||||||||||||||||:::||||||||||||||||||||::::::
Db   809 ATGTTCCTGGGACATCCGGGAATTATAAGGAATGAAGAAGAGTGAAGGATTAGCGGAGA 864
```

RESULT 10
CA775004
LOCUS   CA775004   765 bp   mRNA   linear EST 03-DEC-2002
DEFINITION ip11g06.y1 HR85 islet Homo sapiens cDNA clone IMAGE:6216755 5'
            similar to TR:060523 O60523 RIBONUCLEASE H TYPE II. [2] TR:060857
            ;, mRNA sequence.
ACCESSION   CA775004
VERSION   CA775004.1 GI:26012477
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 765)
AUTHORS   Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
            Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S.,
            Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A.,
            Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J.,

```
Qy  121  SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal  140
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  400  AGCGTGGAGCCGGCGCCTCCAGTTAGCAGAGACACGTTTCCTACATGGAGACTTCGTC  459

Qy  141  ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle  160
              ||||||||||||||||||||||||||||||||||||||:::||||||||||||||
Db  460  GTCGTCTACACCTGATGGCTGCTGCTCCAGTAATGGCCGTAGAAGGCCGCAGCAGGAATC  519

Qy  161  GlyValTyrTrpGlyHisProLeuAsnValGlyIleArgLeuProGlyArgGln  180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  520  GGCGTTTACTGGGGGCCAGGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGCGCAG  579

Qy  181  ThrAsnGlnArgAlaGlyIleHisAlaAlaCysLysGlnGlnAlaLysThrGln  200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  580  ACAAACCAAAGAGCGGAAATTCATCGAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA  639

Qy  201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  640  AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC  699

Qy  221  TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyTrpLysGluValIleAsn  240
              ||||||||||||||||||||||||||||||||||||||||||||:::||
Db  700  TGGGTTCAAGTTGGAAGAAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCCAC  759

Qy  241  LysGlu  242
              ||||||
Db  760  AAAGAG  765
```

RESULT 11
```
BQ921568
LOCUS       BQ921568    AGENCOURT_8922346 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6470333
                         935 bp    mRNA    linear    EST 20-AUG-2002
DEFINITION  5', mRNA sequence.
ACCESSION   BQ921568
VERSION     BQ921568.1  GI:22336599
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 935)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM4000  row: o  column: 06
            High quality sequence stop: 563.
FEATURES            Location/Qualifiers
     source         1..935
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:6470333"
                    /tissue_type="leiomyosarcoma"
                    /lab_host="DH10B (phage-resistant)"
                    /clone_lib="NIH_MGC_71"
                    /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
                    Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                    Average insert size 2.1 kb.  "
ORIGIN
```

Alignment Scores:
```
Pred. No.:                      935
Score:              6.33e-110
                    1281.00     Matches:    255
Percent Similarity:  93.17%     Conservative:  4
```

TITLE   Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R.,
            Williams,T., Jackson,Y. and Bowers,Y.
            Endocrine Pancreas Consortium
JOURNAL   Unpublished (2000)
COMMENT   Other ESTs: ip11g06.x1
            Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
            Endocrine Pancreas Consortium
            Harvard University, Howard Hughes Medical Institute
            Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
            MA 02138
            Tel: 617-495-1812
            Fax: 617-495-8557
            Email: dmelton@biohp.harvard.edu
            Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
            Washington University Genome Sequencing Center For information on
            obtaining a clone please contact: Dr. Hiroshi Inoue
            (hinoue@im.wustl.edu)
            Seq primer: -40RP from Gibco
            High quality sequence stop: 469.
FEATURES   Location/Qualifiers
     source   1..765
                    /organism="Homo sapiens"
                    /mol_type="mRNA"
                    /db_xref="taxon:9606"
                    /clone="IMAGE:6216755"
                    /tissue_type="Purified pancreatic islet"
                    /lab_host="DH10B"
                    /clone_lib="HR85 islet"
                    /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                    NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                    Size-selected on agarose gel. Average insert size ~1kb. 5'
                    XhoI site was destroyed after directional cloning.
                    Amplified once. Contact information: Hiroshi Inoue, MD,
                    Metabolism Div. (Alan Permutt Lab), Washington University
                    School of Medicine, Box 8127, 660 South Euclid Ave., St.
                    Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                    314-362-1916, Fax: 314-747-2692."
ORIGIN

Alignment Scores:
Pred. No.:   765
Score: 1.02e-110
                    1288.00   Matches: 237
Percent Similarity: 99.17%   Conservative: 3
Best Local Similarity: 97.93%   Mismatches: 2
Query Match: 83.31%   Indels: 0
Db: 14   Gaps: 0
```

US-10-054-313-1 (1-286) x CA775004 (1-765)
```
Qy   1   MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaLeuProCysArgArg 20
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 40   ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCCTGGCGGCGCTTGCCCTGCCGCCGC 99

Qy 21   GlySerArgArgGlyPheGlyMetPheTyrAlaValArgArgGlyArgGlyThrGlyValPhe 40
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db 100 GGCTCTCGCGGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGACCGGGGGTCTTT 159

Qy 41   LeuThrTrpAsnGluCysArgAlaGlnValAspArgArgPheProAlaAlaArgPheLysLys 60
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db 160 CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTTCCTGCTGCCAGATTTAAGAAG 219

Qy 61   PheAlaThrGluAspArgGlnValAlaTrpAlaArgLysSerAlaSerProGluValSer 80
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 220 TTTGCCACAGAGGATGAGGCCCTGGGCCTTTGTCAGGAAAATCTGCAAGCCCGGAAGTTCA 279

Qy 81   GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyAlaLysArgLeuArg 100
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db 280 GAAGGGCATGAGAAATCAACATGCACAAGAATCGGAGCGAAAGCCAGCAAGCGACTCCGT 339

Qy 101 GluProLeuAspAlaGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
              |||||||||||||||||||||||||||:::||||||||||||||||||
Db 340 GAGCCACTGGATGGAGATGGACATGGAAGCCGCAGAGCCGTATGCAAAGCACATGAAGCCG 399
```

```
Best Local Similarity:    91.73%            Mismatches:        11
Query Match:              82.86%            Indels:            8
DB:                                         Gaps:              0

US-10-054-313-1 (1-286) x BQ921568 (1-935)

Qy     1  MetSerTrpLeuPheLeuAlaHisArgValAlaAlaAlaLeuAlaProCysArgArg    20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    98  ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGCCTTGGCCCGCCTGCCCGCCGC   157

Qy    21  GlySerArgGlyPheGlyMetPheTyrAlaValAlaArgArgGlyArgGlyValPhe    40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   158  GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGGCAGCAGGGGCGAGGGGGTCTTT   217

Qy    41  LeuThrTrpAsnGluCysArgGlnValAspAspArgPheProAlaArgPheLysLys    60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   218  CTGACCTGGAATGAGTGCAGACAGGTGGACGACCGGTTCCTGCTGCCAGATTTAAGAAG   277

Qy    61  PheAlaThrGluAspGluAlaTrpAlaThrProAlaPheValArgLysSerAlaSerProGluValSer    80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   278  TTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAGTCAGCAAGTCCCGGAAGTTTCA   337

Qy    81  GluGlyHisGluAsnGlnHisGlyGlnValSerGluAlaLysProGlyTyrLysArgLeuArg   100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   338  GAAGGGCATGAAAAATCAACATGGACAAGGCGGAAGCCAAGCCAGCGACTCCCGT   397

Qy   101  GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro   120
          ||||||||||||||||||||||||||||||||||||||:::||||||||||||||||||
Db   398  GAGCCACTGGATGGAGATGGACATGAGAGCGCAGAGCCGTATGCAAAGCACATGAAGCCG   457

Qy   121  SerValGluProProValSerArgAspThrPheSerTyrMetGlyAspPheVal   140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   458  AGGGTGGAGCCGCCGGGCCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC   517

Qy   141  ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgProAlaGlyIle   160
          ||||||||||||||||||||||||||||:||::|||||||||||||||
Db   518  GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGGCCGAGCCGGAATC   577

Qy   161  GlyValTyrTrpProGlyProHisProLeuAsnValGlyIleArgLeuProGlyArgGln   180
          |||||||||||||||||||||||||||:||||||||||:|||||||||
Db   578  GGGGTTTACTGGGGGCCGGGCCCCATCCTTTANATGTAGGCATTAGACTTCCTGGGCGGCAG   637

Qy   181  ThrAsnGlnArgAlaGlulleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln   200
          ||||||||||||||||||||||||||:||||||||||||||||||||
Db   638  ACAAACCAAAGAGCAGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA   697

Qy   201  AsnlleAsn-LysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlylleThrArg   220
          |||||||| |||||||||:|||||||||::|||||||||||:::||||||
Db   698  AACATCAATAAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAAATGGTATAACTAA   757

Qy   220  nTrpValGlnGlyTyrThrLys-LysAsnGlyTrpLysTyrSerThrSerAla--GlyLysGluValArl   239
          ||||   |||||   ::||||| |||:||
Db   758  CTGGGNGTTCAGGTTGGAAGANAAAATGGGGTGGAGACAGTGCANGGAAAAGAGGGTGAT   817

Qy   239  eAsnLysGluAspPheValAlaLeu-GluArgLeuThrGlnGlyMet-AspIleGlnTrp   258
          ||||||||||||||:||||||||| ::||||||||:|||||| |||:|||||
Db   818  CAACAAAGAGACTTTGTGGCACTGGAGAGCTTAACCCAGGGAATGGGACATCAGNGGN   877

Qy   259  MetHis-ValProGlyHis-SerGlyPheIleGlyAsn   270
          ||||| |||||||| :||||||:||
Db   878  ATGCATGGTTCCTGGGCATTTCGGGATTAATAGGCCAT   915
```

```
REFERENCE       Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
                1 (bases 1 to 779)
AUTHORS         NIH-MGC http://mgc.nci.nih.gov/.
TITLE           National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL         Unpublished (1999)
COMMENT         Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: DCTD/DTP
                cDNA Library Preparation: Ling Hong/Rubin Laboratory
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
                Plate: LLCM759 row: i column: 11
                High quality sequence stop: 746.

FEATURES             Location/Qualifiers
     source          1..779
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3928402"
                     /tissue_type="adenocarcinoma cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_9"
                     /note="Organ: ovary; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G). Size-selected >500bp for average
                     insert size 1.8kb. Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies)."

ORIGIN

Alignment Scores:
Pred. No.:                3.83e-106        Length:           779
Score:                    1239.50          Matches:          238
Percent Similarity:       95.63%           Conservative:     3
Best Local Similarity:    94.44%           Mismatches:       8
Query Match:              80.17%           Indels:           4
DB:                                        Gaps:             1

US-10-054-313-1 (1-286) x BE746490 (1-779)

Qy    18  CysArgArgGlySerArgGlyPheGlyMetPheTyrAlaValAlaArgArgGlyArgGlyLysThr    37
          |||||||   :  :|||||||||||||||||||||||||||||||||||||||
Db     2  TGCCGCCGG---TCTCGCGGGTTCGGGATGTTCTATGCCGTGGCAGCAAGGGGCCGAGGAGGGCCGCAAGACC    58

Qy    38  GlyValPheLeuThrTrpAsnGluCysArgGlnValAspAspArgPheProAlaAlaArg    57
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    59  GGGGTCTTTCTGACCTGGAATGAGTGCAGACAGGTGCAGACCAGGTGGACCCGGTTCCTGCTGCCAGA   118

Qy    58  PheLysLysPheAlaThrGluAspGluAlaThrProAlaPheValArgLysSerAlaSerPro    77
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   119  TTTAAGAAGTTTGCCACAGAGGATGAGGCCTGGGCCTTTGTCAGGAAGTCAGCAAGTCCG   178

Qy    78  GluValSerGluGlyHisGluAsnGlnHisGlyGlnValSerGluAlaLysProGlyLys    97
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   179  GAAGTTTCAGAAGGGCATGAAAAATCAACATGGACAAGGCAGAAGCCAAGCCAGCAAG   238

Qy    98  ArgLeuArgGluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHis   117
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   239  CGACTCCCGTGAGCCACTGGATGGAGATGGACATGAGAGCGCAGAGCCGTATGCAAAGCAC   298

Qy   118  MetLysProSerValGluProValAlaProProValSerArgAspThrPheSerTyrMetGly   137
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   299  ATGAAGCCGAGCGTGGAGCCGGTGGCCCGCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGA   358

Qy   138  AspPheValValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArg   157
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   359  GACTTCGTCGTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCGA   418

Qy   158  AlaGlyIleGlyValTyrTrpProGlyProHisProLeuAsnValGlyIleArgLeuPro   177
```

```
RESULT 12
BE746490
LOCUS           BE746490                   779 bp    mRNA        linear    EST 15-SEP-2000
DEFINITION      6015T9445F1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3928402 5',
                mRNA sequence.
ACCESSION       BE746490
VERSION         BE746490.1      GI:10160482
KEYWORDS        EST.
SOURCE          Homo sapiens (human)
  ORGANISM      Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
Db   419  GCAGGAATCGGCGGTTACTGGGGGGCCAGGCCATCCTTTAAATGAGGCATTAGACTTCCT  478
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   178  GlyArgGlnThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAla  197
Db   479  GGGCGGCGACAACCAAAGAGACGGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCA  538

Qy   198  LysThrGlnAsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGly  217
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   539  AAGACTCAAAACATCAATAAACTGGTTCTGTATACAGACAGTAGTGTTTACGATAAATGGT  598

Qy   218  IleThrAsn-TrpValGlnGlyTrpLysLysAsnGlyTrpLysThrSerAlaGlyLysGl  237
          |||||||||  ||||||||||||||||||||||||||||||||
Db   599  ATAACTAACGTGGGTTCAAGTTGGAAGAAAATGGGTGGAAGACAACGTCAGGGAAAGA  658

Qy   237  uValIleAsnLysGluAspPheValAlaLeuGlu-ArgLeuThrGlnGlyMetAspIleG  257
          ||||||||||||||||||||||   :::|||||||||||||||
Db   659  GGTGATCACAGAAGAGGACTTTGTTGGCACTGGAGAAGGCTTACCAGGGGATGGACATTC  718

Qy   257  lnTrpMetHisValProGlyHisSerGlyPhe  267
          ||||||||||   ||||||||||||
Db   719  AGTGGATGCT-GTTCCTGGTCATTCGGGATTT  749
```

RESULT 13
BU595053
```
LOCUS       BU595053                 818 bp    mRNA    linear   EST 20-SEP-2002
DEFINITION  AGENCOURT_8844002 NIH_MGC_142 Homo sapiens cDNA clone IMAGE:6450776
            5', mRNA sequence.
ACCESSION   BU595053
VERSION     BU595053.1  GI:23246812
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 818)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: NCI
            cDNA Library Preparation: Michael Brownstein Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2621  row: p  column: 09
            High quality sequence stop: 533.
FEATURES             Location/Qualifiers
     source          1..818
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6450776"
                     /tissue_type="mixed (pool of 40 RNAs)"
                     /lab_host="DH10B (T1-phage-resistant)"
                     /clone_lib="NIH_MGC_142"
                     /note="Vector: pDNR-LIB; Site_1: Sfi1 (ggccattatggcc);
                     Site_2: SfiI (ggccgcctcggcc);Double-stranded cDNA was
                     prepared from a pool of 40 cell line polyA+ RNAs (bladder
                     - 2%, blood - 33.4%, brain - 5.6%, breast - 12.5%, colon -
                     4%, connective tissue - 1.4%, eye - 1%, intestine - 2.6%,
                     kidney - 2.2%, liver - 5.7%, lung - 10.8%, NK-cell - 5.2%,
                     ovary - 4%, pharynx - 2.5%, prostate - 4.3%, salivary
                     gland - 1.3%, and skin - 2.3%). 5' and 3' adaptors were
                     used in cloning as follows:
                     5'-AAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGGG-3' and
                     5'-ATTCTAGAGGCCGAGGCGGCCGACATG-dT(30)NN-3'. Full-length
                     enriched library was constructed using the Clontech
                     Creator SMART kit and size-selected to contain the >0.5 kb
                     size fraction (other fractions present in NIH_MGC_141).
                     Library created in the laboratory of M. Brownstein (NIMH,
                     NIH). Note: this is a NIH_MGC Library."

ORIGIN
```

Alignment Scores:
```
Pred. No.:          4.02e-105
Score:              1229.00          Length:         818
Percent Similarity:  96.37%         Matches:        236
Best Local Similarity: 95.16%       Conservative:     3
Query Match:         79.50%         Mismatches:       5
DB:                                 Indels:           4
                                    Gaps:             0
```

US-10-054-313-1 (1-286) x BU595053 (1-818)

```
Qy     1  MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaLeuAlaLeuProCysArgArg   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    67  ATGAGCTGGCTTCTCGTTCTCCTGGCCCACAGAGTCGCCTTGGCCCTGGCCCTGCCGCCGC  126

Qy    21  GlySerArgGlyPheGlyMetPheTyrAlaValArgGlyLysArgGlyLysThrGlyValPhe   40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   127  GGCTCTCGCGGGTTCGGGATGTTCTATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT  186

Qy    41  LeuThrTrpAsnGluCysArgAlaGlnIleValAspArgPheProAlaAlaArgPheLysLys   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   187  CTGACCTGGAATGAGTGCAGAGCACAGGTGGACCGGTTCCTTGCTGCCAGATTTAAGAAG  246

Qy    61  PheAlaThrGluAspGluAlaTrpAlaPheValArgLysSerAlaSerProGluValSer   80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   247  TTTGCCACAGAGGATGAGGCCTGGCCTTTCGTCAGAAATCTGCAAGCCCGGAAGTTTCA  306

Qy    81  GluGlyHisGluAsnGlnHisGlyGlnGluSerGluAlaLysProGlyLysArgLeuArg  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   307  GAAGGGCATGAAAAATCAACATGCAGCAAGAATCGGAGGCGAAAGCCAAGCCAGGCCGT  366

Qy   101  GluProLeuAspGlyAspHisGluSerAlaGlnLeuProTyrAlaLysHisMetLysPro  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   367  GAGCCACTGGATCGAGATGACATCAGAAAGCGCAGAGCCGTATCGCAAAGCACATGAAGCCG  426

Qy   121  SerValGluProAlaProProValSerArgArgAspThrPheSerTyrMetGlyAspPheVal  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   427  AGCGTGGAGCCCGCGCCGCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC  486

Qy   141  ValValTyrThrAspGlyCysCysSerSerAsnGlyArgArgLysProArgAlaGlyIle  160
          |||||||||||||||||||||||||||||:::|||||||||||||||||||||
Db   487  GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCGAGCAGGAATC  546

Qy   161  GlyValTyrTrpGlyTyrProGlyTyrHisProLeuAsnValGlyTyrIleArgProGlyArgGln  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   547  GGGGTTTACTGGGGCTACCCATCCTTAAATGTAGGCATTAGACTTCCTGGGCGGCCAG  606

Qy   181  ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysAlaIleGluGlnAlaLysThrGln  200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   607  ACAAACCAAAGAGACGGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAA  666

Qy   201  AsnIleAsnLysLeuValLeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsn  220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   667  AACATCAATAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATGGTATAACTAAC  726

Qy   221  TrpVal-GlnGlyTrpLysLysAsnGlyTrpLys-ThrSerAla-GlyLysGluValIle  239
          |||||  |||||||||||||||||||||   ||||||||  |||||||||
Db   727  TGGGTTCCAGGGTTGGAAGAAAATGGGTGGAAGANAACGNAGTGCAGGGGAAAGAGGTGATC  786

Qy   240  Asn-LysGluAspPhe  244
          :::  |||||||||
Db   787  CACAAAAGAGGACTTT  802
```

RESULT 14
BM451915
```
LOCUS       BM451915                 916 bp    mRNA    linear   EST 05-FEB-2002
DEFINITION  AGENCOURT_6379386 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5492781
            5', mRNA sequence.
ACCESSION   BM451915
VERSION     BM451915.1  GI:18500955
```

```
Qy  141 ValValTyrThrAspGlyCysCysSerAsnGlyArgArgLysProArgAlaGlyIle 160
        ||||||||||||||||||||||||||||: :||||||||||||||||||||||||
Db  495 GTCGTCTACACTGATGGCTGCTGCTCCAGTAATGGGCGTAGAAGGCCGCAGGAATC 554

Qy  161 GlyValTyrTrpGlyHisProLeuAsnValGlyIleArgLeuProGlyTyrArgGln 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  555 GGGGTTTACTGGGGGCCAGGCCATCCTTAAATGTAGGCATTAGACTTCCTGGGCGCAG 614

Qy  181 ThrAsnGlnArgAlaGluIleHisAlaAlaCysLysValAlaIleGluGlnAlaLysThrGln 200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  615 ACAAACCANAGAGCGGAAATTCATGCAGCCTGCAAAGCCATTGAACAAGCAAAGACTCAN 674

Qy  201 AsnIleAsn---LysLeuValLeuValTyrThrAspSerMetPheThrIleAsnGlyIleThrAs 220
        |||||||      ||:||||||||||||||||||||||||||||||||||||||||||||
Db  675 AACATCCATAAAACTGGTTCTGTATACAGACAGTATGTTTACGATAAATAATGTATAACTAA 734

Qy  220 nTrp-ValGlnGly-TrpLysTyrAsn-GlyTrpLysTyrSerAla-GlyLysGluValI 239
        ||||  |||||||| |||||||||||| |||||||||||||||||| |||||||||||||
Db  735 CTGGGGTTCCAGGTTTGGAAGAAAAATGGGGTGGAAGACCAGTGCAGGGAAAGAGGTGA 794

Qy  239 leAsn---LysGluAspPheValAlaLeuGluArgLeuThrGlnGlyMetAspIleGlnT 258
        |||:     ::: ||||||||||||||||||||||||||||||||||||---------GGAT
Db  795 TTCACCAAAGAAGGACTTTTGTGCACTGGAAGGCTCTACCCAGGG---------GGAT 845

Qy  258 rpMetHis------------ValProGlyHisSerGly---PheIleGlyAsnGluG 272
        ||||                |||||||||||||::|   |||||||||||||
Db  846 GGACCTTTCCCCGGGATGCTTGTTTCCCGGGGCCATCCGGGCATTTTTAGGCCAATGAAA 905

Qy  272 lu 272
        ::
Db  906 AA 907
```

```
RESULT 15
BF984400
LOCUS       BF984400               869 bp    mRNA    linear   EST 23-JAN-2001
DEFINITION  602307923F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4399301 5',
            mRNA sequence.
ACCESSION   BF984400
VERSION     BF984400.1  GI:12387212
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 869)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10102 row: n  column: 06
            High quality sequence stop: 691.
FEATURES    Location/Qualifiers
     source          1..869
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4399301"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC 88"
                     /note="Organ: small_intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
```

```
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 916)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12115 row: g  column: 22
            High quality sequence stop: 626.
FEATURES    Location/Qualifiers
     source          1..916
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5492781"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC 67"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT;
                     Average insert size 1.75 kb. Library constructed by Life
                     Technologies."

ORIGIN
Alignment Scores:
Pred. No.:
Score:                    2.18e-104
                          1222.00
Percent Similarity:       89.08%            Length:          916
Best Local Similarity:    86.97%            Matches:         247
Query Match:              79.04%            Conservative:    6
DB:                       12                Mismatches:      16
                                            Indels:          15
                                            Gaps:            4

US-10-054-313-1 (1-286) x BM451915 (1-916)

Qy    1 MetSerTrpLeuLeuPheLeuAlaHisArgValAlaLeuAlaAlaAlaLeuProCysArgArg 20
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   75 ATGAGCTGGCTTCTGTTCCTGGCCCACAGAGTCGGCTTGGCCGCCGCCCTTGCCCTGCCCGGC 134

Qy   21 GlySerArgGlyPheGlyMetPheTyrAlaValArgArgValLysThrGlyValValPhe 40
        |||||||||||||||||||||||||||||||||||||||||||||||
Db  135 GGCTCTCGCGGGTTCGGGATGTTCATGCCGTGAGGAGGGGCCGCAAGACCGGGGTCTTT 194

Qy   41 LeuThrTrpAsnGluCysArgAlaGlnIleValAlaAspArgPheProAlaAlaArgPheLysLys 60
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  195 CTGACCTGGAATGAGTGCAGAGCACAGATCGTTGCCGACCGGTTCCTGCCGTCCCAGATTAAGAAG 254

Qy   61 PheAlaThrGluAspGluAspTrpAlaTrpAlaPheValArgLysSerAlaSerProGluValSer 80
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  255 TTTGCCACAGAGGATGAGGACCTGGGCCCTTTGTCAGGAAAATCTGCAAGCCCCGAAGTTCA 314

Qy   81 GluGlyHisGluAsnGlnHisGlyHisGlySerGluAlaLysProGlyLysArgGluLeuArg 100
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db  315 GAAGGGCATGAAAATCAACATGGACAAGAAATCCAGGCGAAAGCCAGCAGGCGACTCCGT 374

Qy  101 GluProLeuAspGlyAspGlyHisGluSerAlaGlnProTyrAlaLysHisMetLysPro 120
        |||||||||||||||||||||:::||||||||||||||||||||||||
Db  375 GAGCCACTGGATGGAGATGGACATGAAAGCGCAGCCGTATGCAAAGCACATGAAGCCG 434

Qy  121 SerValGluProAlaProProValSerArgAspThrPheSerTyrMetGlyAspPheVal 140
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db  435 AGCGTGGAGCCCGCGCCCCCTCCAGTTAGCAGAGACACGTTTTCCTACATGGGAGACTTCGTC 494
```

Tue Mar 23 14:48:36 2004

us-10-054-313-1.p2n.rst

ORIGIN

```
Alignment Scores:
Pred. No.:                9.16e-104      Length:           869
Score:                    1215.00        Matches:          231
Percent Similarity:       97.90%         Conservative:     2
Best Local Similarity:    97.06%         Mismatches:       4
Query Match:              78.59%         Indels:           2
DB:                       10             Gaps:             0
```

US-10-054-313-1 (1-286) x BF984400 (1-869)

```
Qy   47  ArgAlaGlnValAspArgPheProAlaAlaArgPheLysPheAlaThrGluAspGlu  66
             ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3  AGAGCACAGGTGACCGGTTTCCTGCTGCCAGATTTAAGAAGTTTGCCACAGAGGATGAG  62

Qy   67  AlaTrpAlaPheValArgLysSerAlaSerProGluValSerGluGlyHisGluAsnGln  86
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63  GCCTGGGCCTTTGTCAGGAAATCTGCAAGCCCCGAAGTTTCAGAAGGGCATGAAAATCAA  122

Qy   87  HisGlyGlnGluSerGluAlaLysProGlyLysArgGluArgGluProLeuAspGlyAsp  106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  123  CATGGACAAGAATCGGAGGGCGAAAGCCAGCAAGCGACTCCGTGAGCCACTGGATGGAGAT  182

Qy  107  GlyHisGluSerAlaGlnProTyrAlaLysHisMetLysProSerValGluProAlaPro  126
             |||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db  183  GGACATGAAAGCGCAGAGCGTATGCAAAGCACATGAAGCCGAGCCGTGGAGCCGGCGCCT  242

Qy  127  ProValSerArgAspThrPheSerTyrMetGlyPheValValValTyrTyrThrAspGly  146
             ||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db  243  CCAGTTAGCAGAGACACGTTTCCTACATGGGAGACTTCGTCGTCGTCTACACTGATGGC  302

Qy  147  CysCysSerSerAsnGlyArgArgLysProArgArgAlaGlyIleGlyValTyrTrpGlyPro  166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  303  TGCTGCTCCAGTAATGGGCGTAGAAGGCCGCAGAGGAATCGGCGTTTACTGGGGGCCG  362

Qy  167  GlyHisProLeuAsnValGlyTyrIleArgGlnThrGlnThrAsnGlnArgAlaGlu  186
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  363  GGCCATCCTTTAAATGTAGGCATTAGACTTCCTGGGCGGCAGACAAACCAAAGAGCCGGAA  422

Qy  187  IleHisAlaAlaCysLysLysAlaIleGluGlnAlaLysThrGlnAsnIleAsnLysLeuVal  206
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  423  ATTCATGCAGCCTGCAAAGCCATTGAACAGCAAAGACTCAAAACATCAATAAACTGGTT  482

Qy  207  LeuTyrThrAspSerMetPheThrIleAsnGlyIleThrAsnTrpValGlnGlyTyrTrpLys  226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  483  CTGTATACAGACAGTATGTTTACGATAAAATGGTATAACTAACTGGGTTCAAGGTTGGAAG  542

Qy  227  LysAsnGlyTrpLysThrSerAlaGlyLysGluValIleAsnLysGluAspPheValAla  246
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  543  AAAAATGGGTGGAAGACAAGTGCAGGGAAAGAGGTGATCAACAAAGAGGACTTTGTGTGCA  602

Qy  247  LeuGluArgLeuThrGlnGlyMetAspIleGlnTrpMetHisValProGlyHisSerGly  266
             ||||||||||||||||||||||||||     ||||||||||||||||||||||||||
Db  603  CTGGAGAGGCTTACCCA-GGGATGGACATTCAGTGGACATGTTCCTGGTCATTCGGGA  661

Qy  267  PheIleGlyAsnGluGluAlaAspArgLeuAlaArgGluGly-AlaLysGln  283
             ||||||||||||||||||        |||||    ||||||||||||
Db  662  TTTATAGGCAATGAAGAAGCTGACAGATTAGGCACAGAGAAGGAGGCTAAACAA  713
```

Search completed: March 23, 2004, 00:10:25
Job time : 2344 secs

US 1046918908P1

Creation date: 06-18-2004
Indexing Officer: SMOHAMMED - SUAD MOHAMMED
Team: 1600PrintWorkingFolder
Dossier: 10469189

Legal Date: 03-12-2004

No.	Doccode	Number of pages
1	WFEE	1

Total number of pages: 1

Remarks:

Order of re-scan issued on ....